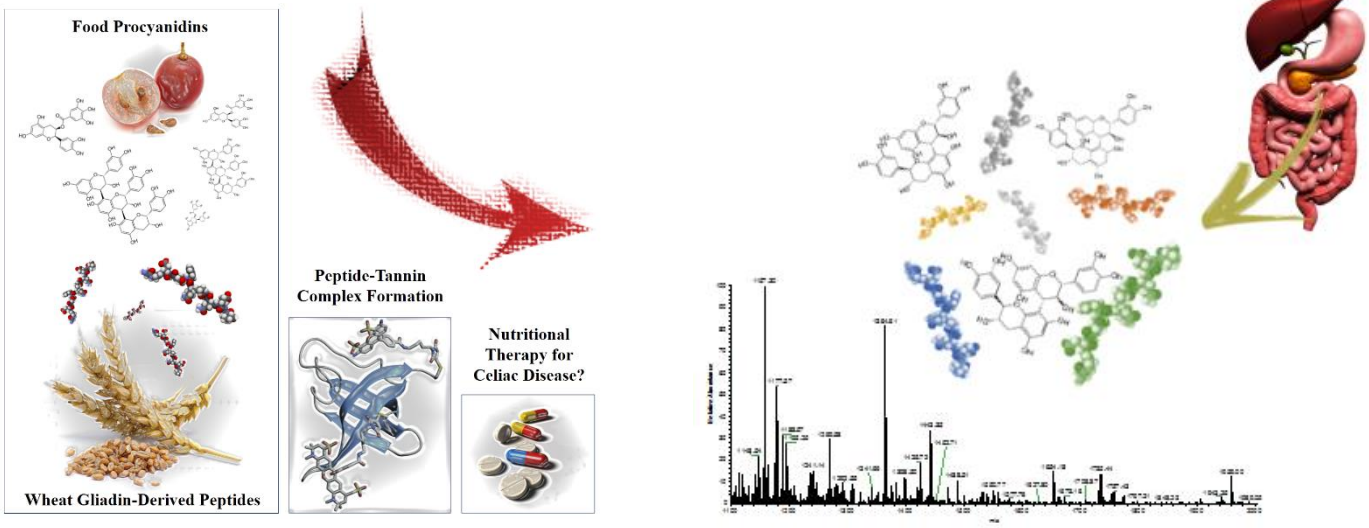
Study of the interaction between food phenolics and Celiac Disease related peptides

Ricardo Jorge Correia Dias

Biochemistry Master's Dissertation presented to

Faculty of Sciences, University of Porto

2015



Study of the interaction between food phenolics and Celiac Disease related peptides

Ricardo Jorge Correia Dias

Master's in Biochemistry

Department of Chemistry and Biochemistry

2015

Supervisor

Dr. Victor Armando Pereira Freitas,

Full Professor, Faculty of Sciences - University of Porto

Co-Supervisor

Dra. Maria Rosa Perez Gregorio,

Pos-Doc Researcher, Faculty of Sciences - University of Porto

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.
O Presidente do Júri,
Porto, _____/_____/_____

ACKNOWLEDGEMENTS

Quase tão difícil quanto foi escrever esta tese atempadamente é conseguir, em tão curto espaço, expressar o mais profundo agradecimento a todos aqueles que, de uma forma ou de outra, tornaram a realização deste trabalho possível. Desta forma, deixo apenas algumas palavras, poucas, mas um sentido e sincero sentimento de reconhecido agracedimento.

Ao meu orientador, *Professor Doutor Victor de Freitas*, agradeço a oportunidade e o previlégio que tive de frequentar este mestrado que muito contribuiu para a minha formação académica e cientifica. Sempre atento e interessado, depositou em mim uma confiança extraordinária e um apoio inigualável. Quero também deixar uma palavra de apreço ao *Professor Doutor Nuno Mateus* pelo seu dinamismo, boa disposição e pela forma como sempre se mostrou disponível a ajudar e a conversar.

Em especial, e por aquilo que foi o meu dia-a-dia, não há palavras que descrevam a enorme gratidão e reconhecimento que reservo à minha co-orientadora, *Rosa*. Foi absolutamente incansável no seu acompanhamento, paciência, auxilio e incentivos constantes que tanto contribuíram para a concretização desta etapa da minha formação académica. Muito obrigado pela amizade e apoio incondicional, pelas criticas e correcções, e pela alegria e entusiasmo contagiantes e enriquecedores que muito estimularam o meu desejo de querer, sempre, saber mais e fazer melhor. Foram vários os desafios que encontramos ao longo destes 7 meses. Cada dia, uma surpresa. Apesar de tudo, com muita vontade e ao lado da pessoa certa, não há nada que não se consiga. Sem ti, nada disto teria sido possível e por isso, esta tese é tanto minha como tua. *Indeed*, uma verdadeira equipa RR!

Gostava de agradecer também ao *Raul* e à *Sílvia* pela amizade, companheirismo, carinho e boa disposição em todos os momentos. Além das palavras de ânimo e confiança que imprimiam sempre que achavam necessário, permitindo que cada dia fosse encarado com especial motivação, foram simplesmente fantásticos pelos conselhos e sugestões, imprescindíveis para o desenvolvimento deste trabalho e para o meu futuro.

Aos meus colegas de laboratório (e de almoço), *Ana Gomes*, *Ana Luisa*, *Elsa*, *Hélder*, *Iva*, *Joana Azevedo*, *Joana Brás*, *Joana Oliveira*, *Mariana*, *Natércia*, *Luís*,

Luisa, *Paulinha* e *Susana*, obrigado pelo apoio, simpatia, conversa e excelente ambiente de trabalho e de convívio proporcionados. Agradeço adicionalmente a todos os elementos do Departamento de Química e Bioquímica da Faculdade de Ciências (FCUP) e ao Centro de Materiais da Universidade do Porto (CEMUP) por me terem fornecido as condições necessárias à elaboração desta tese de mestrado.

Aos meus amigos e família, um enorme obrigado por acreditarem sempre em mim e naquilo que faço e por todos os ensinamentos de vida. Espero que esta jornada, que agora termino, possa, de alguma forma, retribuir e compensar todo o carinho, apoio e dedicação que, constantemente, me oferecem. A eles, dedico todo este trabalho.

ABSTRACT

Celiac Disease (CD) is an autoimmune disorder of the small intestine that occurs in genetically predisposed people of all ages from middle infancy onward. It is caused by an inflammatory reaction towards gluten proteins found in wheat, and similar proteins from other common grains such as barley, rye and oats. Given the high prevalence and lack of therapeutic means to treat it besides a gluten-free diet, the search for drugs and nutraceuticals that can block the initial stages of this chronic disease is a priority. Among the diversity of polyphenols, tannins have been described as the most reactive towards proline-rich proteins, which are structurally similar to gluten peptides responsible for the onset of CD. Therefore, the aim of this work was to verify the ability of different food tannins to interact with gliadin-derived peptides by means of fluorescence quenching and dynamic light scattering experiments. The characterization of soluble tannin-peptide complexes resulting from the association between procyanidin B3 and different gluten peptides was performed through electrospray ionization mass spectrometry (ESI-MS).

Following an *in vitro* digestion of a wheat gliadins raw extract, the resulting peptides were fractionated by semi-preparative HPLC and further characterized by ESI-MS/MS to determine their sequence as well as the proteins by which they derived. The results presented herein show differences in those fractions composition where the presence of peptides containing known CD epitopes was revealed. Using procyanidin B3, procyanidin trimers, procyanidin tetramers and an oligomeric mixture of high molecular weight procyanidins it was demonstrated, for the first time, that an association between those phenolic molecules and the previous peptide mixtures does occur, although in different contexts. Indeed, at the micromolar level it was observed by means of fluorescence assays that the tannins size and structural features is related to their quenching ability as a result of specific interactions or complex formation. So, for the same peptide mixture, the smaller procyanidin (B3) was the weakest quenching molecule because it was the one that provided fewer binding groups. However, in different peptide mixtures, the same polyphenolic molecule could have different binding affinities, which is probably related to the differential amino acid composition of the respective peptides. At the millimolar level by using DLS, it was concluded that the procyanidins reactivity towards different peptide mixtures is mainly dependent on those

peptides size with drastic effects on the dimension of the resulting aggregates. Finally, using a mass spectrometry approach, several soluble B3-peptide complexes were detected and identified. It was observed that the number, size and amount of identified complexes differed among the collected fractions, with some of them involving peptides with varied CD T-cell epitopes. Although a binding selectivity of procyanidin B3 towards peptides containing CD epitopes was not found, all the immunoreactive ones were involved in the formation of different non-covalent complexes.

Overall, this study clearly opens new therapeutical perspectives for celiac disease, by using phenolic compounds as a nutraceutical approach to enhance the return of the full intestinal function in patients who show incomplete recovery in response to a gluten-free diet.

Keywords: Celiac Disease, Polyphenols, T-Cell Epitopes, Tannins, Tannin-Peptide Interactions, Bottom-up Proteomics, Wheat Proteins.

INDEX

ACKNOWLEDGEMENTS I
ABSTRACT III
INDEX VII
LIST OF FIGURES XI
LIST OF TABLES XVII
LIST OF ABBREVIATIONS XIX
OBJECTIVES XXI
INTRODUCTION 1
1.1 CELIAC DISEASE 3
1.1.1 The Gluten Factor 4
1.1.1.1 Wheat Gluten Proteins Nomenclature, Composition and Structure 6
1.1.1.2 Relationship Between Gluten and Celiac Disease 8
1.1.2 The Genetic Factor 9
1.1.2 Other Environmental Factors 9
1.1.4 Pathophysiological Mechanisms of CD 10
1.1.5 Current Therapeutical Options and Need for Alternatives 12
1.2 POLYPHENOLS 14
1.2.1 Tannins 18
1.2.1.1 Condensed Tannins (Proanthocyanidins) 19
1.2.3.2 Hydrolysable Tannins 21
1.2.3 Interaction between Tannins and Proteins 22
1.2.4 Tannin Bioavailability 24
1.2.5 Tannin Biological Effects 26
EXPERIMENTAL 27
2.1 REAGENTS 29

2.2 GRAPE SEED TANNIN ISOLATION 29
2.3 PROCYANIDIN B3, PROCYANIDIN TRIMER T1 AND PROCYANIDIN TETRAMER TT1 SYNTHESIS 30
2.4 CHARACTERIZATION OF WHEAT GLIADINS RAW EXTRACT 30
2.4.1 MALDI-TOF-MS Analysis. 30
2.4.2. Protein Separation by SDS-PAGE. 31
2.4.3. Tryptic digestion. 31
2.4.4. Protein Identification by Mass Spectrometry. 32
2.4.4.1. ESI-MS/MS Analysis. 32
2.4.4.2. Nano-ESI-MS/MS analysis. 32
2.4.4.3. MALDI-TOF-MS Analysis. 33
2.5 SEPARATION AND IDENTIFICATION OF GLIADIN DERIVED PEPTIDES 34
2.5.1 In vitro Digestion of Gliadin Raw Extract. 34
2.5.2 Fractionation of Peptides Obtained after Wheat Gliadin Digestion. 34
2.6 PEPTIDE-TANNIN INTERACTION ASSAYS 35
2.6.1 Fluorescence Quenching Measurements. 35
2.6.2 Dynamic Light Scattering Measurements. 36
2.6.3 Identification of Peptide-B3 Complexes by Mass Spectrometry. 36
2.7 STATISTICAL ANALYSIS 37
RESULTS AND DISCUSSION 39
3.1 GRAPE SEED TANNIN FRACTIONS IDENTIFICATION 41
3.2 CHARACTERIZATION OF WHEAT GLIADINS RAW EXTRACT 42
3.2.1 MALDI-TOF-MS Analysis of Wheat Gliadins Raw Extract. 42
3.3 SEPARATION AND CHARACTERIZATION OF WHEAT GLIADIN PEPTIDES 50
3.4 FLUORESCENCE QUENCHING STUDIES 56
3.5 LIGHT SCATTERING STUDIES 61
3.6 CHARACTERIZATION OF PROCYANIDIN B3-GLIADIN DERIVED PEPTIDE COMPLEXES 64
CONCLUSIONS 73
REFERENCES 77

SUPLEMENTARY INFORMATION 95

S1 IDENTIFICATION OF PROTEINS CONTAINED IN THE RAW EXTRACT BY NANO-ESI-MS/MS 97

S1.1 BSA Tryptic Digestion Analysis 97

S1.2 Wheat Gliadins Raw Extract Protein Identification 99

S1.2.1 Band 1' 99

S1.2.3 Band 2´ 101

S1.2.3 Band 3´ 102

S1.2.4 Band 4´ 104

S1.2.5 Band 5´ 106

S1.2.6 Band 6´ 106

LIST OF FIGURES

Page

Figure 1 – Classification scheme of gluten proteins according to Osborne and Shewry. The latter subdivision is represented for wheat prolamins and glutelins (i.e. gliadins and glutenins respectively). 5

Figure 2 – Molecular mechanisms underlying CD pathogenesis: interaction of gluten with environmental, immune and genetic factors. Adaptado de Sollid *et al.* 2013. 12

Figure 3 – Chemical structures of the four major classes of polyphenols according to the number of phenolic rings that they contain and the structural elements that bind these rings to one another. Some examples of food products in which each class is abundant are also represented. 16

Figure 4 – Chemical structures of the six major subclasses of flavonoids according to the type of heterocyclic C ring involved and its substitution pattern. Some examples of food products in which each class is abundant are also represented. 18

Figure 5 – Structures of simple dimeric procyanidins to demonstrate different linkage types in condensed tannins. 20

Figure 6 – General structure of proanthocyanidins 20

Figure 7 – Examples of structures of hydrolysable tannins: pentagalloylglucose (gallotannin) (A), punicalagin (ellagitannin) (B). 21

Figure 8 – Scheme of the interaction between condensed tannins and proteins: main driving forces (hydrogen bonds and hydrophobic interactions) between phenolic rings (cross-linkers) of tannins and the 22

amide groups and apolar side chains of amino acids such as proline. Adapted from Santos-Buelga *et al.* 2008.

Figure 9 – Proposed three-step model for the interaction between randomly coiled proteins and multidentate polyphenols. Adapted from Jobstl *et al.* 2004. 24

Figure 10 – ESI-MS spectra of oligomeric procyanidins from grape seeds. To highlight and differentiate the high molecular weight procyanidins, the original spectrum was enlarged between 1000 and 2000 m/z. 41

Figure 11 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract (10-55 kDa) using both sinapinic acid (SA) and 2,5-dihydroxybenzoic acid (2,5-DHB) as matrices. 44

Figure 12 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract using different analyte/SA ratios: 1:1 (orange color), 1:2 (red color) and 2:1 (green color). 45

Figure 13 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract. 46

Figure 14 – SDS-PAGE of the wheat gliadins raw extract using BSA as an internal control. Molecular weight (MW) standards (left) were used as mass references (kDa). 47

Figure 15 – MALDI-TOF-MS analysis of tryptic digests. 49

Figure 16 – Identification of proteins in the wheat gliadins extract by MALDI-TOF-MS, SDS-PAGE and Nano-ESI-MS/MS. 50

Figure 17 – Chromatographic profile, acquired by HPLC semi-preparative, of the simulated in vitro digestion of the wheat gliadins raw sample. Each peptide mixture was collected in different time intervals, as represented in the chromatogram. 51

Figure 18 – UV and Total Ion Current Chromatograms (TIC) of the different gliadin-derived peptidic fractions. As it shows, several differences were found between fractions as they differed mainly in their polarity. Given the low peptide concentration needed to verify their identity by mass spectrometry, the signal obtained in UV was very poor. However, the sequenciation, after MS/MS analysis, of the peptides contained in each fraction could be reported since the concentration used was high enough to be detected by the mass spectrometer. 52

Figure 19 – Number (n) and size distribution (average number of amino acid residues per peptide) of gliadin-derived peptides, obtained after *in vitro* digestion of the wheat gliadins crude extract and further fractionation. It displays the median and Tukey test of the obtained values. 54

Figure 20 – Number and precursor protein distribution of all sequenced peptides in each collected fraction. 55

Figure 21 – Fluorescence emission spectra (at λ_{em} = 290 nm) of *Pep Mix4* (0.02 mg/mL) in the presence of increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FII of oligomeric procyanidins. Each curve represents a triplicate assay after correction for polyphenol fluorescence. 57

Figure 22 – Fluorescence emission spectra (at λ_{em} = 290 nm) of *Pep Mix6* (0.02 mg/mL) in the presence of increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FII of oligomeric procyanidins. Each curve represents a triplicate assay after correction for polyphenol fluorescence. 58

Figure 23 – Stern-Volmer plots describing tryptophan quenching of *Pep Mix4* (A) and *Pep Mix6* (B) by increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FII of oligomeric procyanidins. The fluorescence emission intensity was recorded at λ_{ex} 290 nm, and the λ_{em} maximum occurred at 365 nm. 59

Figure 24 – Differences in aggregation size of tannin-PepMix4 complexes depending on the peptide concentration (A, 0.02 mg/mL and B 0.2 mg/mL) and 62

as the tannin tested (procyanidin B3 and procyanidin tetramer TT1).

Figure 25 – Changes in the aggregate size of all the seven collected peptide mixtures (0.2 mg/mL) by increasing concentrations of FII containing oligomeric procyanidins with a high degree of polymerization. 63

Figure 26 – ESI-MS spectra with all detected B3-CRP32 complexes. *1*. $[B3+H^+]^+$, *2*. $[B3 + Na^+]^+$, *3*. $[B3+K^+]^+$, *4*. $[CRP32 + 4H^+]^{+4}$, *5*. $[CRP32 + Na^+ + 3H^+]^{+4}$, *6*. $[CRP32 + K^+ + 3H^+]^{+4}$, *7*. $[CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *8*. $[CRP32 + 2Na^+ + K^+ + H^+]^{+4}$, *9*. $[CRP32 + Na^+ + 2K^+ + H^+]^{+4}$, *10*. $[B3\text{-}CRP32 + 4H^+]^{+4}$, *11*. $[B3\text{-}CRP32 + Na^+ + 3H^+]^{+4}$, *12*. $[B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *13*. $[B3\text{-}CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *14*. $[B3\text{-}CRP32 + 2Na^+ + K^+ + H^+]^{+4}$, *15*. $[B3\text{-}CRP32 + Na^+ + 2K^+ + H^+]^{+4}$, *16*. $[CRP32 + 3H^+]^{+3}$, *17*. $[CRP32 + Na^+ + 2H^+]^{+3}$, *18*. $[CRP32 + K^+ + 2H^+]^{+3}$, *19*. $[CRP32 + Na^+ + K^+ + H^+]^{+3}$, *20*. $[B3\text{-}CRP32 + 4H^+ - NH_4^+]^{+3}$, *21*. $[B3\text{-}CRP32 + 3H^+]^{+3}$, *22*. $[B3\text{-}CRP32 + Na^+ + 2H^+]^{+3}$, *23*. $[B3\text{-}CRP32 + K^+ + 2H^+]^{+3}$, *24.* $[B3\text{-}CRP32 + Na^+ + K^+ + H^+]^{+3}$, *25*. $[B3\text{-}CRP32 + 2Na^+ + K^+]^{+3}$, *26*. $[2B3\text{-}CRP32 - NH_4^+ + 4H^+]^{+3}$, *27*. $[2B3\text{-}CRP32 + 3H^+]^{+3}$, *28*. $[2B3\text{-}CRP32 + Na^+ + 2H^+]^{+3}$, *29*. $[2B3\text{-}CRP32 + K^+ + 2H^+]^{+3}$, *30.* $[2B3\text{-}CRP32 + Na^+ + K^+ + H^+]^{+3}$, *31*. $[2B3\text{-}CRP32 + 2Na^+ + K^+]^{+3}$, *32*. $[4B3\text{-}CRP32 + 4H^+]^{+4}$, *33*. $[4B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *34*. $[4B3\text{-}CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *35*. $[4B3\text{-}CRP32 + 2K^+ + 2H^+]^{+4}$, *36*. $[3B3\text{-}CRP32 + 4H^+]^{+4}$, *37*. $[3B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *B3* - Procyanidin dimer B3, *CRP32* - QLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF. 65

Figure 27 – ESI-MS spectra of *Pep Mix6* and *B3-Pep Mix6* mixture assay. The most pronounced complexes (only 14 out of 20) are highlighted by red numbers and further characterized in terms of their composition. *1.* $[DPLGAL + B3 + H^+]^{+1}$, *2.* $[LQPQNPSQQQPQEQVPL + B3 + 2H^+]^{+2}$, *3.* $[QQSSYQVL + 3B3 + K^+ + H^+]^{+2}$ *or* $[QLEMMTSI + 3B3 + K^+ + H^+]^{+2}$ *or* $[SSSPQQLGQGQPRY + 2B3 + K^+ + H^+]^{+2}$ *or* $[TQQPQQPFPQQPQQPFPQ + B3 + 2H^+]^{+2}$ *or* $[MQLQPFPQPQLPYPQPQL + B3 + 2H^+]^{+2}$, *4.* $[QQSSYQVL + 3B3 + 2K^+]^{+2}$ *or* $[QLEMMTSI + 3B3 + 2K^+]^{+2}$ *or* $[SSSPQQLGQGQPRY + 2B3 + 2K^+]^{+2}$ *or* $[TQQPQQPFPQQPQQPFPQ + B3 + K^+ + H^+]^{+2}$ *or* $[MQLQPFPQPQLPYPQPQL + B3 + K^+ + H^+]^{+2}$, *5.* $[GQVQWPQ + B3 + H^+]^{+1}$, *6.*Unknown, *7.* Unknown, *8.* $[QQSSYQVL + B3 + H^+]^{+1}$ *or* $[QLEMMTSI + B3 + H^+]^{+1}$, *9.* Unknown, *10.* $[LQPQNPSQQQPQEQVPL + 2B3 + Na^+ + H^+]^{+2}$, *11.* $[SSSPQQLGQGQPRY +$ 67

3B3 + K^+ + $H^+]^{+2}$ *or* [TQQPQQPFPQQPQQPFPQ + 2B3 + $2H^+]^{+2}$ *or* [MQLQPFPQPQLPYPQPQL + 2B3 + $2H^+]^{+2}$, *12.* [TQQPQQPFPQQPQQPFPQ + 2B3 + Na^+ + $H^+]^{+2}$, *13.* [SSSPQQLGQGQPRY + 3B3 + $2K^+]^{+2}$ *or* [TQQPQQPFPQQPQQPFPQ + 2B3 + K^+ + $H^+]^{+2}$ *or* [MQLQPFPQPQLPYPQPQL + 2B3 + K^+ + $H^+]^{+2}$, *14.* Unknown.

Figure 28 – Qualitative parameters related to the identification of B3-gliadin derived peptide complexes: number of unknown and identified non-covalent complexes (with or without epitopes). 69

Figure 29 – Molecular weight (Da) of B3-gliadin derived peptide complexes found in the different peptidic fraction assays. The average molecular weight of each fraction non-covalent complexes is displayed by the + symbol. 70

Figure S1 – Nano-LC-MS chromatogram of BSA tryptic digest. 97

Figure S2 – Nano-LC-MS spectra of PEG. 99

Figure S3 – Nano-LC-MS chromatogram of band 1' tryptic digest. 100

Figure S4 – Nano-LC-MS chromatogram of band 2' tryptic digest. 101

Figure S5 – Nano-LC-MS chromatogram of band 3' tryptic digest. 103

Figure S6 – Nano-LC-MS chromatogram of band 4' tryptic digest. 104

Figure S7 – Nano-LC-MS chromatogram of band 5' tryptic digest. 106

Figure S8 – Nano-LC-MS chromatogram of band 6' tryptic digest. 107

LIST OF TABLES

Page

Table 1 – Characterization of wheat gluten protein types. Adapted from Wieser 2007 6

Table 2 – Chromatographic conditions for the identification of gluten proteins. 32

Table 3 – ESI-MS/MS conditions for the identification of gluten proteins. 32

Table 4 – Nano-ESI-MS/MS conditions for the identification of gluten proteins. 33

Table 5 – ESI-MS analysis of the five oligomeric procyanidin fractions. The mean molecular weight of the fractions was determined based in the relative abundance of each flavanol present. 41

Table 6 – Identification of some protein components present in the wheat gliadins raw extract. 48

Table 7 – Identification, by LC-ESI-MS/MS, of peptides generated after enzymatic hydrolysis of the wheat gliadins raw extract. For each collected peptide mixture, different peptide colors and asterisks are indicative of a specific precursor protein. The peptides containing known epitopes that are recognized by intestinal T cells of CD patients are underlined and highlighted in bold. 53

Table 8 – Sequenced peptides containing CD epitopes (identified in different colors) as well as their precursor protein code and name. 56

Table 9 – Stern-Volmer Quenching Constants (Ksv) for the interaction 60

between both Pep Mix4 and Pep Mix6 and procyanidins with increasing degree of polymerization (B3, trimer T1, tetramer TT1 and FV of oligomeric procyanidins). Values with different letters (a-e) are significantly different (P < 0.05).

Table 10 – Bimolecular Quenching Constants (kq) for the interaction between both Pep Mix4 and Pep Mix6 and procyanidins with increasing degree of polymerization (B3, trimer T1, tetramer TT1 and FV of oligomeric procyanidins). Values with different letters (a-e) are significantly different (P < 0.05). 61

Table 11 – Identification of B3-peptide non-covalent complexes by ESI-MS. The molecular weight (Da) as well as their hypothesized identification is shown. The complexes involving peptides with CD epitopes are highlighted in red. 68

Table S1 – BSA sequenced peptides by Nano-LC-ESI-MS/MS. 98

Table S2 – Protein identification according to the peptides sequenced using Proteome Discoverer 1.4 software. 109

LIST OF ABBREVIATIONS

A-PAGE - Acid-polyacrylamide gel electrophoresis;
APC - Antigen presenting cell;
BSA - Bovine serum albumin;
CD - Celiac Disease;
CHCA - α-cyano-4-hidroxycinnamic acid;
CID - Collision-induced dissociation;
CRP - Celiac reactive peptide;
DC - Dendritic cell;
DHB - 2,5-dihydroxybenzoic acid;
DLS - Dynamic light scattering;
ESI-MS - Electrospray ionization mass spectrometry;
FRET - Förster resonance energy transfer;
HCl - Hydrochloric acid;
HLA - Human leukocyte antigen;
HMW-GS - High-molecular-weight glutenin subunit;
HPLC - High-performance liquid chromatography;
IE-CTLs - Intraepithelial cytotoxic T lymphocytes;
IFN-γ - Interferon gamma;
IL-10 - Interleukin 10;
IL-15 - Interleukin 15;
K_q - Bimolecular quenching constant;
K_{sv} - Stern-Volmer constant;
LC - Liquid chromatography;
LMW-GS - Low-molecular-weight glutenin subunit;
MALDI-TOF-MS - Matrix-assisted laser desorption/ionization - time-of-flight mass spectrometer;
MHC - Major histocompatibility complex;
MIC-A - Major histocompatibility complex class I chain-related A;
MLN - Mesenteric lymph node;
MS - Mass spectrometry;
MS/MS - Tandem mass spectrometry;

NCE - Normalized collision energy

NK-G2D - Killer cell lectin-like receptor subfamily K, member 1;

NMR - Nuclear magnetic resonance;

PEG - Polietilenglicol;

PEP - Prolyl endopeptidase;

PTFE - Polytetrafluoroethylene;

RSD - Relative standard deviation;

SA - Sinapinic (or sinapic) acid;

SDS-PAGE - Sodium dodecyl sulfate-polyacrylamide gel electrophoresis;

TFA - Trifluoroacetic acid;

Th - T helper;

TIC - Total Ion Current;

TNF-α - Tumor necrosis factor alpha;

tTG2 - Tissue Transglutaminase 2.

UHPLC - Ultra Performance Liquid Chromatography.

OBJECTIVES

Despite the high prevalence and severe symptoms, presently, the only accepted treatment for Celiac Disease (CD) involves a strict dietary abstinence of wheat gluten and similar proteins from rye, barley and certain oat varieties. Among the high diversity of polyphenols, tannins have been described as the most reactive towards proline-rich proteins, which are structurally similar to celiac reactive peptides (CRPs). As these bioactive compounds present low intestinal absorption and suffer reduced metabolism in the human digestive system, they remain in the small intestine for extended periods of time, a feature that should allow their interaction with gluten proteins and/or CRPs through essentially hydrophobic and hydrogen bonding. As vegetal tannins present a good potential as therapeutic agents for blocking the development of CD from both a nutraceutical and a pharmacologic point of view, the main goals of this study are:

1. Characterize a commercially available wheat gliadins raw extract by bottom-up proteomics;
2. Fractionate and identify, by LC-ESI-MS/MS, the peptides obtained after a gastric and pancreatic *in vitro* digestion of the previous wheat proteins;
3. Determine, by means of fluorescence quenching, the binding affinity and interaction mechanism between increasingly polymerized food tannins and the different peptide fractions;
4. Evaluate, using dynamic light scattering experiments, the reactivity of each collected peptide fraction towards different tannin procyanidins, by measuring the size distribution of the resulting aggregates;
5. Follow the binding of a common food tannin (procyanidin B3) and different wheat-derived peptides by ESI-MS and further characterize the different non-covalent complexes that coexist in solution in terms of their relative composition/ stoichiometries.

INTRODUCTION

1.1 CELIAC DISEASE

Celiac Disease (CD) (also known as Celiac Sprue) is an acquired and prevalent food hypersensitivity disorder triggered by the ingestion of gluten proteins from widely prevalent food sources such as wheat, rye, barley and to lower extent, certain oat varieties [1, 2]. Chronic inflammation of the small intestinal mucosa typically results in villous atrophy, crypt hyperplasia, dense lymphocytic infiltration and a variety of clinical symptoms that differ according to the age group [3, 4]. These include fatigue, diarrhea, abdominal distension, malabsorption and extra-intestinal symptoms such as osteoporosis and neurological disorders [5, 6]. It is documented that, in certain points of time, the disease may not be correlated with obvious clinical signs, as it can persist in a latent or silent status [7]. If untreated, it is associated with increased morbidity and mortality [8-10]. The diagnosis of CD, straightforward in most cases, is usually established on the basis of examination of duodenal biopsies, serological testing and a positive response to the only effective treatment currently accepted: a gluten-free diet [5, 11].

Epidemiological studies reports that CD affects approximately 1% of the world's population [12, 13]. The disease is now becoming widely recognized not only throughout Europe and in countries populated by people of European ancestry but also in the Middle East, Asia, South America and North Africa. Even if in most affected people CD remains undiagnosed, its incidence rates are currently growing due to the increasing use of serological screening that leads to the diagnostic of milder cases [14, 15].

Current knowledge about the pathogenesis of CD implicates a complex interplay between environmental and many genetic factors. In fact, among the high diversity of chronic inflammatory conditions in which the identification of the underlying pathogenic mechanisms is difficult, CD emerges as a particularly instructive model disorder since:

- The environmental factor that precipitates the disease is known (dietary gluten) and can be withdrawn and re-introduced in a controlled manner;
- The HLA molecules that confer predisposition to disease have been identified (HLA-DQ2 for the majority of patients and HLA-DQ8 for a minority);
- Access to the affected organ (the small intestine) is simple, enabling detailed *in situ* studies and isolation of disease relevant cell populations.

The conjugation of these three unique factors allowed to verify that CD develops as a result of an abnormal innate and CD4+ T-cell-initiated immune response to gluten proteins [16].

1.1.1 The Gluten Factor

In genetically predisposed individuals, CD is precipitated by the ingestion of gluten, a term that collectively refers to the enormous family of complex and heterogeneous storage proteins found in the endosperm of cereal grain proteins [17]. Despite cereal grains only contain about 8-20% of protein, they ultimately provide a considerable proportion of the daily nutritional requirements of protein needed by humans and livestock worldwide [18]. Apart from their nutritional value, cereal seed proteins also own interesting functional properties, a feature that allows them to play a crucial role in the manufacturing and processing of a great diversity of food products - bread, noodles, pasta, cookies, cakes, pastries, among others [19-21].

Over the years, storage cereal grain proteins have been extensively studied and several attempts have been made in order to further subdivide and classify the proteins encompassed in this family according to their characteristics, amino acid sequences, location within the grain, solubility and relationship within other grains. Osborne, in 1908, was the one who developed the first (and still used) classification system of gluten proteins based on their solubility in various solvents. Accordingly, proteins were divided into four major types (*Fig. 1*): albumins (soluble in water and dilute buffers), globulins (soluble in dilute salt solutions), prolamins (soluble in 70-90% ethanol) and glutelins (soluble in dilute acid or alkali) [22]. A different classification scheme was subsequently proposed in which prolamins and glutelins could be grouped together based on their amino acid composition into sulphur-poor, sulphur-rich and high molecular weight (HMW) proteins (*Fig. 1*) [19].

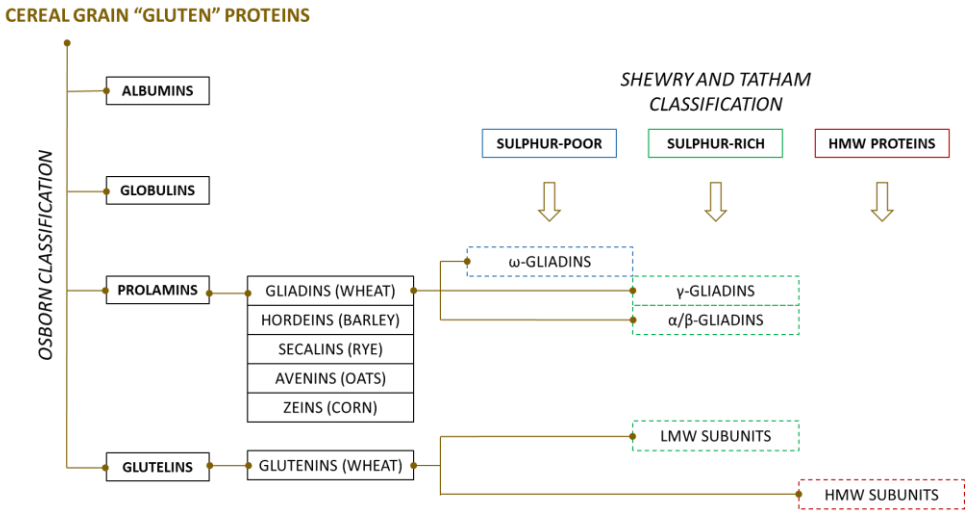


Fig. 1 – Classification scheme of gluten proteins according to Osborne and Shewry. The latter subdivision is represented for wheat prolamins and glutelins (i.e. gliadins and glutenins respectively).

Nowadays, the definition of gluten has evolved and is generally used to describe the water-insoluble seed storage proteins found in the *Triticeae* tribe of the grass (*Gramineae*) family. It is therefore the rubbery mass that is left when wheat, rye or barley flour is washed with water to remove starch granules, non-starch carbohydrates and water-soluble constituents [23, 24]. Gluten is considered to be one of the most complex protein families found in nature. Indeed, it classically consists of hundreds of aqueous alcohol-soluble prolamins and alcohol-insoluble glutelins, present either as monomers or, linked by inter-chain disulphide bonds, as polymers. Both fractions are important contributors to the rheological properties and baking quality of dough by conferring water absorption capacity, cohesivity, viscosity and elasticity. Although these proteins differ in terms of their amino acid composition, they are all characterized by high contents of glutamine and proline residues and by low levels of charged amino acids such as glutamic and aspartic acid. The molecular weights of those native proteins range from about 28 kDa to more than 10 million kDa [24]. The properties of the various types of gluten proteins are summarized in *Table 1*.

Table 1 – Characterization of wheat gluten protein types. Adapted from Wieser 2007.

TYPE	MOLECULAR WEIGHT (KDA)	PROPORTIONS[a]	PARTIAL AMINO ACID COMPOSITION (%)				
			GLUTAMINE	PROLINE	PHENYLALANINE	TYROSINE	GLYCINE
ω5-GLIADINS	49-55	3-6	56	20	9	1	1
ω1,2-GLIADINS	39-44	4-7	44	26	8	1	1
α/β-GLIADINS	28-35	28-33	37	16	4	3	2
γ-GLIADINS	31-35	23-31	35	17	5	1	3
x-HMW-GS	83-88	4-9	37	13	0	6	19
y-HMW-GS	67-74	3-4	36	11	0	5	18
LMW-GS	32-39	19-25	38	13	4	1	3

In recent years, many studies have been conducted focusing on characterizing the various classes of gluten proteins [25-27]. Although numerous amino acid sequences have been determined (mostly deduced from nucleotide sequences) for representatives of the prolamin and glutelin sub-groups of wheat, rye and barley, gaps still exist in the knowledge base for structure, composition and functionality of all members of this enormous protein family.

1.1.1.1 Wheat Gluten Proteins Nomenclature, Composition and Structure

The principal toxic components of wheat gluten, the environmental stimuli responsible for both initiation and maintenance of the disease process have been identified and belong to a family of closely related proline and glutamine-rich proteins designated gliadins [10, 28]. These consist of a polymorphic mixture of essentially monomeric proteins with most found to be alcohol soluble. The gliadins molecular weight ranges from 28 to 55 kDa and initially they were divided into α-, β-, γ- and ω-gliadins, on the basis of their mobility at low pH in acid polyacrylamide gel electrophoresis (A-PAGE). Later biochemical and genetic studies, however, suggested that the electrophoretic mobility does not always reflect the protein relationships and that α and β-gliadins fall into one group. Meanwhile, gliadins were grouped into four different classes (α/β-, ω5-, ω1,2- and γ-gliadins) being this division based on their different amino acid sequences and molecular weights (*Table 1*) [24, 29].

Within each protein subtype, small structural differences are due to substitutions, deletions and insertions of single amino acid residues. ω-gliadins are the molecular components that have the higher molecular weights (according to some literature, they can reach up to 75 kDa) and are characterized by the highest percentage of proline,

glutamine and phenylalanine amino acid residues (*Table 1*). These proteins consists almost entirely of repetitive sequences of 8 to 10 amino acids rich in proline and glutamine. Since most ω-gliadins lack cysteine residues, they are not able to form intra- or inter-molecular disulphide crosslinks. α/β- and γ-gliadins, on the other hand, have similar molecular weights (~28-35 kDa) and proportions of glutamine and proline residues much lower than those of ω-gliadins. Each of both types contain unique repetitive sequences of 11 (for α/β-) and 7 (for γ-) amino acids in their N-terminal domains and homologous, non-repetitive, C-terminal domains with 6 (for α/β-) and 8 (for γ-) cysteines that are involved in the establishment of intra-chain crosslinks. Other structural domains are also present in these proteins and are made up of long repetitive sequences of individual glutamines as well as sequences rich in proline and glutamine. Studies on the secondary structure of α/β-, ω- and γ-gliadins describe their N-terminal domain as consisting by β-turns whereas the C-terminal domain contains considerable proportions of α-helices and β-sheets. In those studies, the secondary structures of gliadins were found to be very sensitive to environmental conditions such as solvent, hydration, and temperature [30]. Although the distribution of total gliadin proteins among the different classes is strongly dependent on wheat genotype and growing conditions, it can be generalized that α/β- and γ-gliadins are the major components while the ω-gliadins occur in much lower proportions in wheat varieties. Hydrated gliadins have little elasticity and are less cohesive than glutenins. They mainly contribute to the viscosity and extensibility of dough system [24, 31-33].

Glutenins are the polymeric proteins of wheat gluten. It comprises aggregated proteins linked together by inter-chain disulphide bonds. They have a varying size ranging from about 500 kDa to more than 10 million kDa thus making them the largest and most complex protein polymers in the plant kingdom. Contrary to gliadins, glutenins appear to be largely responsible for gluten elasticity. Indeed, dough properties and baking performance are strongly determined by those proteins molecular weight distribution [34, 35].

Despite being considered the alcohol insoluble components of wheat gluten in their native/polymeric form, after disulphide bonds reduction with reducing agents, the resulting monomeric glutenin subunits do become soluble in aqueous alcohol similar to gliadins. Based on primary structure, glutenin subunits have been divided into two classes: the high-molecular-weight subunits (HMW-GS) and low-molecular-weight subunits (LMW-GS). From those, LMW-GS are the predominant protein type, constituting about 20% of the total wheat gluten proteins. They are related to α/β- and γ-gliadins in terms of their molecular weight range (32 to 35 kDa), amino acid composition and structural domains. While their N-terminal domain is characterized by

short proline- and glutamine-rich repetitive sequences, the C-terminal one has six cysteine residues which are proposed to be linked by intra-chain disulphide bonds. For steric reasons, two additional and unique cysteine residues are not able to form an intra-chain disulphide bond. Instead, inter-chain disulphide bonds involving those cysteines of different gluten proteins are generated [24, 31, 33, 36, 37].

HMW-GS belong to the minor components of the wheat gluten protein family as this fraction represents about 10% overall. The HMW glutenins from each wheat variety consists of 3 to five HMW-GS which contain 3 structural domains: a non-repetitive N-terminal domain, a repetitive central domain and a C-terminal domain. The N- and C-terminal domains are characterized by the frequent occurrence of charged amino acids and the presence of most cysteine residues. The middle domain contains repetitive sequences of 4 to 6 amino acids. According to the structure of their N- and central domains, the HMW-GS can be grouped into two different subtypes: the x- (83-88 kDa) and the y-type (67-74 kDa). It has been demonstrated that among HMW-GS, whose quantity strongly influences the dough properties, the contribution of the x-type subunits is much more important than that of the y-type. Limited information is available about the secondary structure of HMW-GS. Studies focusing on the repetitive central domain indicated that it consists of β-reverse turns which are predicted to be overlapping, thus forming a loose spiral. This feature appears to be decisive to gluten elasticity although it is dependent on several environmental parameters affecting protein structure and properties (presence of salt, CO_2, starch, etc). The N- and C-terminal domains were proposed to have globular domains containing α-helices [24, 33, 38].

Although covalent bonds consisting of inter- and intra-chain disulphide bonds between and within gluten proteins are a prevalent structural characteristic of this complex and heterogeneous family proteins, it should be highlighted that the gluten network is also superimposed by non-covalent bonds (hydrogen, ionic and hydrophobic bonds). Collectively, these types of molecular interactions are responsible for the various functionalities of gluten proteins in manufacturing processes. Similar groups of prolamin and glutelin proteins are also found in barley, rye, corn and oat grains. Even if in general they all have a similar structure, at a detailed level, considerable differences are noticed [19, 24, 39, 40].

1.1.1.2 Relationship Between Gluten and Celiac Disease

Usually, after food ingestion, the respective protein content is gradually hydrolyzed into increasingly smaller peptides and free amino acids by host proteases, before they can be transported across the intestinal mucosa. Nevertheless, due to the unusually high proline content of gluten proteins from wheat (and related cereals) and because human proteases are unable to hydrolyze amide bonds when they are adjacent to conformationally constrained proline residues, then proline-rich peptides become protected from proteolysis by gastric, pancreatic and intestinal brush border membrane enzymes. Hence, relatively stable and small gluten-derived peptides (~4 to 50 amino acids in length) are allowed to reach the intestinal lamina propria at a higher concentration than those obtained from other food proteins where they can exert toxic effects in genetically susceptible individuals [10, 29, 41]. Although other toxicological effects of exposure to gluten have also been described [42, 43], an inflammatory reaction in the small intestine mediated by both innate and adaptive immune systems to these incompletely metabolized gluten peptides is currently the best characterized one [4, 44].

1.1.2 The Genetic Factor

Genetic background appears to be a major risk factor for CD since it will not develop in an individual unless they have inherited the necessary genetic factors that are part of the immunological response to gluten [15]. In that way, it is well established that CD phenotypic expression is strongly associated with specific human leukocyte antigen (HLA) class II alleles that map to the DQ locus of the major histocompatibility complex (MHC): HLA-DQ2 (DQA1*05/DQB1*02) and HLA-DQ8 (DQA1*0301/DQB1*0302) [16, 45]. Despite their unequivocal involvement in conferring susceptibility to CD, many people, most of whom do not have the disease, carry these alleles. This feature suggests that their presence is necessary but not sufficient for disease development [5], thus making HLA-typing successful as a diagnostic tool to identify genetic risk but not for defining CD. Several other genetic factors must also be considered such as those that influence the innate and adaptive immune system and the physical integrity of the intestinal barrier [46]. Further testing of those non-HLA candidates is awaited.

1.1.2 Other Environmental Factors

In addition to gluten, several other environmental factors that have an important role in the development of CD have been suggested by recent epidemiological studies [47]. Among these, the occurrence of certain gastrointestinal infections, such as rotaviral infections, have been reported to increase the risk for the onset of CD in genetically predisposed children [48]. Causing a transient rise in small-bowel permeability, bacterial infections could lead to an up-regulation and release of tissue transglutaminase that, in turn might enhance gluten immunogenicity [49].

A different line of evidences suggests a possible implication of the infant-feeding practices in the disease pathogenesis [50]. In that way, it has been demonstrated that introduction of small amounts of dietary gluten while infants were still being breastfed may be a more important protective factor in preventing or minimizing the disease risk in childhood [51-53].

1.1.4 Pathophysiological Mechanisms of CD

Interplay between innate and adaptive immune responses to ingested gluten is involved in CD characteristic tissue damage and villous atrophy of the small intestinal mucosa (*Fig. 2*) [54, 55]. In the epithelium, proline-rich gluten fragments that survived processing by luminal and brush-border proteases can directly affect intestinal cell structure and function by different, strictly related, molecular mechanisms such as intracellular tight junction dysfunctions, rearrangements of actin cytoskeleton, modulation of gene expression, altered cell differentiation and apoptosis. Most of these effects are mediated by increased oxidative stress in enterocytes which reflects in an impairment of the epithelial barrier and increased permeability [4]. As a consequence, several CD biologically active peptides are allowed to pass through the intestinal epithelial layer, by both transcellular and paracellular routes, and interact with antigen-presenting cells (APCs) in the lamina propria. APCs mature in response to interleukin-15 (IL-15) and type I interferon (IFN) produced by stressed enterocytes, and acquire pro-inflammatory properties. After migration to the draining Peyer's patch or mesenteric lymph node (MLN), mature DCs present gluten peptides that have undergone deamidation by the enzyme tissue transglutaminase 2 (tTG2), to induce the activation of gluten-specific HLA-DQ2 or HLA-DQ8-restricted $CD4^+$ T cells [2, 56-58]. Transglutaminase 2 (tTG2), which in the intestinal mucosa in located mainly extracellularly in the sub-epithelial region (also found in the brush-border), is a Ca^{2+}-dependent enzyme that is responsible for the post-translational modification of proteins by transamidation or deamidation of specific polypeptide-bound glutamines. The

transamidation (cross-linking) reaction involves a primary amine (i.e. lysine) while the deamidation reaction (conversion of glutamine residues to negatively charged glutamic acids) involves water [59]. The propensity for deamidation compared with transamidation is increased by lowering the pH and by increasing the concentration ratio of glutamine substrates to primary amines [16, 60, 61]. This indicates that the deamidation of gluten peptides in the small intestine might occur in a slightly acid environment. In CD patients, the deamidation of glutamine residues that remain in glutamine/proline-rich gluten peptides leads to an enhanced immunogenicity of those [56, 62]. Once activated, gluten-reactive $CD4^+$ T cells proliferate and produce mainly Th1-type cytokines, particularly IFN-γ (interferon gamma) and TNF-α. (tumor necrosis factor alpha) [7, 63]. The secretion of Th1 cytokines activates the release of matrix metalloproteinases and other tissue-damaging mediators that induce crypt hyperplasia and villous injury [4, 64].

TG2-specific auto-reactive B cells can internalize TG2-gluten peptide complexes and consequently present these peptides on HLA-DQ2 or HLA-DQ8 at their surface. Similarly, gluten-specific B cells are able to bind and present deamidated gluten peptides in a conventional manner. The gluten-specific $CD4^+$ T cells provide help to both auto-reactive TG2-specific and gluten-specific B cells, which differentiate into antibody-producing plasma cells [57]. Activated gluten-specific CD4+ T cells also provide signals (that remain to be fully defined) to pre-activated epithelial cells, which up-regulate the expression of IL-15 and non-classical MHC class I molecules such as HLA-E and MIC-A (major histocompatibility-complex class I chain-related A). Consequently, intraepithelial cytotoxic T lymphocytes (IE-CTLs) expressing the natural-killer-cell marker receptor NK-G2D become activated and kill epithelial cells on the basis of the recognition of stress signals [5, 57, 65-67]. Overall, this leads to a chronic feedback of the inflammatory process and a progressive destruction of the intestinal villous structure as long as CD toxic gluten oligopeptides are present in the intestinal lumen.

To date, several DQ2- and DQ8-restricted T cell epitopes from disease-associated grains have been identified. In wheat, those epitopes are derived from various gluten proteins including the α-, γ- and ω-gliadin subtypes and low-molecular-weight (LMW) glutenins [28, 62, 68, 69]. Among the high diversity of the already characterized CD epitopes, there are some of them that stand out because of their extraordinary immunoreactivity as is the case of the immunodominant 33-mer peptide [10].

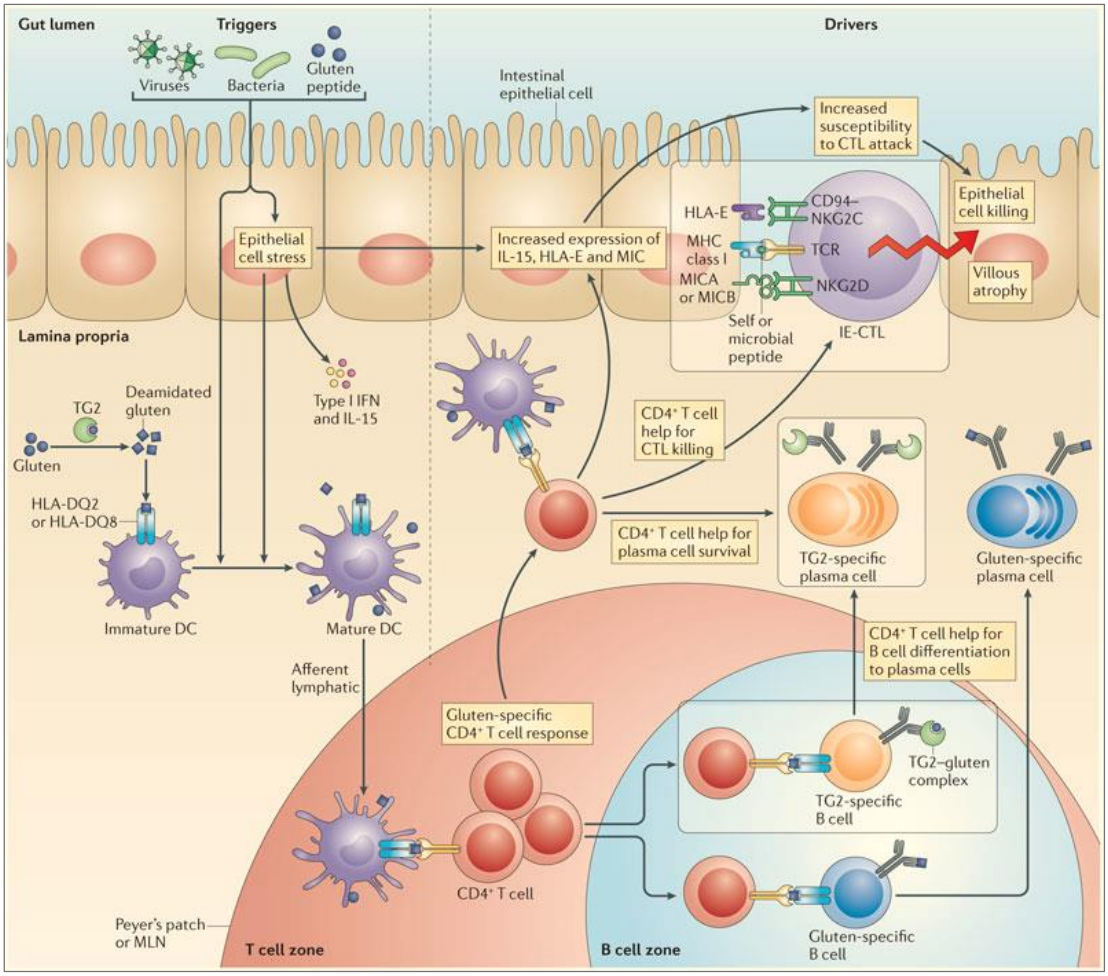


Fig. 2 – Molecular mechanisms underlying CD pathogenesis: interaction of gluten with environmental, immune and genetic factors. Adaptado de Sollid *et al*. 2013.

1.1.5 Current Therapeutical Options and Need for Alternatives

Despite the high prevalence and severe symptoms, presently, the only accepted treatment for CD involves a strict, lifelong adherence to a gluten-free diet [2, 5, 7, 11]. Elimination of gluten usually induces clinical improvements within days or weeks, thought histological recovery takes months or even years, especially in adults whose mucosal recovery may be incomplete [70]. In sporadic cases, children tolerate re-introduction of a normal diet after a long-term clinical and histological response [71]. However, a complete avoidance of gluten is not easily achieved. It takes time, motivation and patience to become accustomed to such a diet. On the other hand, gluten-"free" products (containing less than 20 parts per million as set by the U.S. Food and Drug Administration) are not widely available and are usually more expensive than their gluten-containing counterparts [2, 5]. Unfortunately, there also exist many hidden

sources of gluten as a consequence of unlabeled ingredients and from cross-contamination in manufacturing processes and equipment cleaning. Taken together, dietary compliance is frequently imperfect in a large fraction of patients, thus restraining successful outcomes [2, 5, 72]. Therefore, there is an urgent need to develop safe and effective therapeutical alternatives to enhance the return of full intestinal function in patients who show incomplete recovery in response to a gluten-free diet. Ideally, it would also allow moderate quantities of gluten to be tolerated [2].

Our understanding of the molecular and cellular bases of CD has made huge progress in recent years. The knowledge of which gluten epitopes are recognized by intestinal T cells should facilitate the methods by which gluten-free foods are assessed and improved. Additionally, our increasing insights into the disease mechanisms should uncover new targets for therapy and consequently benefit patients. Therefore, there is currently a great interest in the development of non-dietary therapeutical approaches that might either replace of supplement the rough gluten-free diet. Already, there are some attractive possibilities that can potentially interfere with the activation of $CD4^+$ gluten-specific T cells and so, be an effective way to control the disease [2, 16, 73, 74]. These include:

- Production of CD-associated grain seeds that are devoid of T-cell epitopes, either by traditional breeding programs or by using transgenic technology [75];
- Oral enzyme supplementation using recombinant prolyl endopeptidases (PEPs) that, in contrast to human gastrointestinal proteases, can readily cleave immunostimulatory proline-rich gluten peptides [41, 76-78]. Although there is a human PEP, this is expressed only in the cytosol and is therefore unlikely to have a physiological role in the digestive destruction of gluten peptides [79];
- Development of drugs capable of inhibiting tTG2 activity either directly or indirectly [58, 80]. In the latter case, as it has been proposed that the deamidation of gluten peptides may occur in endosomes, then drugs that affect the acidification of APC endosomes in the small intestine might be suitable for the treatment of CD. However, since tTG2 is involved in many physiological processes such as programmed cell death, then interfering with its transamidation and deamidation activity might result in unacceptable side-effects [81, 82]. The existence of T-cell epitopes that do not need to be modified by TG2 for recognition to occur make this approach unfeasible;
- Silencing of gluten-reactive T cells. This feature could be achieved by peptide-based immunotherapy, a promising strategy that is currently being evaluated with encouraging results for both allergies and autoimmune diseases. This type

of immunotherapy aims at the immune tolerance recovery, generating a shift of the immune response and concomitantly the induction of regulatory T-cells [62, 83]. Alternatively, one could try to silence gluten-specific T cells directly using soluble dimers of HLA-peptide complexes thus inducing their apoptosis as a result of inappropriate stimulation [84]. However the increasing number of characterized gluten epitopes with the ability to trigger immunological reactions in CD patients would complicate this approach;

- Cytokine therapy using interleukin 10 (IL-10) to counterbalance the characteristic CD Th1 immune response or antibodies capable of neutralizing IFN-γ and IL-15 [2, 85, 86];
- Selective inhibition of leucocyte adhesion molecules to prevent their migration into inflamed tissues. As clinical candidates there is the integrin-α4 and integrin-α4β7 antagonists [2];
- Blocking the peptide-binding sites of HLA-DQ2 and HLA-DQ8 molecules. Here, the challenge will be to find an efficient way to target and block those DQ molecules as they are continuously synthetized by APCs [2, 5, 16, 87];
- Etc.

A comprehensive understanding of the processes that lead to CD pathology places it in a unique position with prospects of developing specific immunotherapy. Nevertheless, despite the outcoming alternatives, new therapeutical modalities will have to prevail over the current gluten-free diet with regards to safety.

1.2 POLYPHENOLS

Widely distributed in almost all plant foods and beverages, polyphenols are a group of secondary metabolites that are characterized by the presence of more than one hydroxyl group attached to at least one phenolic unit per molecule. Among the high diversity of those, they might be associated with various carbohydrates and organic acids or even with one another [88]. To date, many structurally different polyphenols have been identified in higher and edible plants and it is believed that there are many others that remain unknown, especially those available in lower quantities [89].

In plants, these compounds have functions that range from pigmentation to growth, including defense against ultraviolet radiation, predators and aggression by

microorganism [90, 91]. In foodstuffs, polyphenols are directly related to several organoleptic characteristics such as flavor, color, aroma and oxidative stability. Through the formation of complexes with salivary proteins, complex polyphenols are responsible for the astringency perception of many fruits and beverages and for the bitterness of chocolate [92-94]. Often perceived as a negative attribute, in some highly and worldwide consumed beverages like wine, tea or beers, a balanced level of astringency is required for their quality [89].

Overall the last years, dietary phenolic compounds have been a strong focus of research because of their biologically significant anticarcinogenic [95-97], antimutagenic [98, 99], antioxidant [95, 100, 101], antiallergenic [102-104], anti-inflammatory [105, 106], anticoagulant [107], antitrombotic [108], antiulcer [109], antiatherogenic [110] and antimicrobial functions [111-113]. Indeed, according to some epidemiological studies, quite a lot of healthy effects have been attributed to the consumption of plant polyphenols as they provide a significant protection against the development of several chronic diseases including cardiovascular and neurodegenerative diseases, cancer, diabetes, osteoporosis, infections, aging, asthma, etc [93, 114]. Recently, this versatile compounds proved to be also effective in the inhibition of HIV (Human Immunodeficiency Virus 1) [115] as well of HSV-1 (Herpes Simplex Virus 1) [116]. Besides these biological activities, some adverse nutritional and toxicological effects have also been reported for dietary polyphenols due to their ability to complex and inhibit digestive enzymes (proteases, glycosidases and lipases), reduce the digestibility of ingested food proteins, depression of growth and acute hepatotoxicity [117-119]. In particular, the aptitude of tannins to bind proteins in a specific and selective [120, 121] manner may further increase their potential applications in diverse knowledge fields as is the case of toxicology were tannin-protein complexes have been used as snake venom antidotes [122, 123]. Nevertheless, it has to be highlighted that the health effects of polyphenols is highly dependent on the nature, amount consumed and on their bioavailability. In that way, not all polyphenols are absorbed with the same efficacy. Additionally, they are extensively metabolized by digestive and hepatic enzymes and by the intestinal microflora being then eliminated in urine and bile. Therefore, depending on the modifications of which dietary polyphenols are subjected during their metabolic processing, their biological activities within target tissues might be very affected [124-126].

In general, phenolic compounds may be divided into four different classes according to the number of phenolic rings that they contain and the structural elements that bind these rings to one another [88]. Distinctions are thus made between phenolic acids (including benzoic and cinnamic acid derivatives), flavonoids, stilbenes and lignans

(*Fig. 3*). One of the most common phenolic acids is caffeic acid, present in many fruits and vegetables. It is most often esterified with quinic acid to form chlorogenic acid, which is the major phenolic compound in coffee [127].

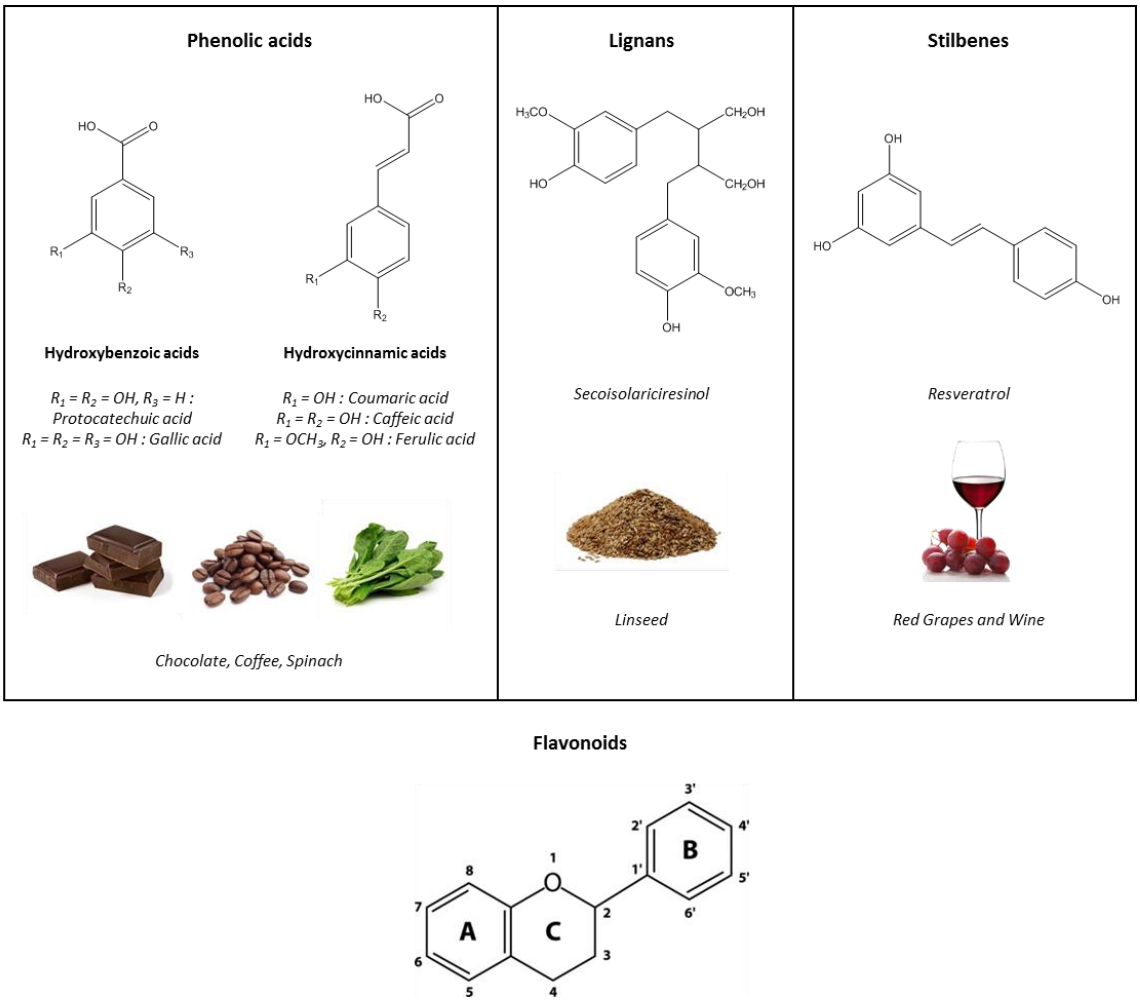


Fig. 3 – Chemical structures of the four major classes of polyphenols according to the number of phenolic rings that they contain and the structural elements that bind these rings to one another. Some examples of food products in which each class is abundant are also represented.

In addition of being the most important class of phenolic compounds found in foodstuffs of plant origin, the flavonoids group is also the most structurally diversified one. At present, more than 4000 unique flavonoids have been identified and the number is still growing. They are a large family of phytonutrients that provide the more vibrant, brilliant colours in nature, including most of the blue, purple and emerald green tones found in flowers, leaves, fruits and vegetables. Moreover, most of the yellow, orange and red colours that are not carotenoids belong to the flavonoid family [128,

129]. In our diet, flavonoids are found primarily in fruits and vegetables. Significant amounts of flavonoids are also found in teas and wines.

The basic structure of flavonoids consists of a typical C6-C3-C6 skeleton (*Fig. 3, bottom*) where two aromatic rings (A and B) are bound together by three carbon atoms that form an oxygenated heterocycle (C ring). According to the type of the heterocycle involved and its substitution pattern, the flavonoids group may be itself divided into six subclasses: flavonols, flavones, isoflavones, flavanones, anthocyanidins and flavanols (*Fig. 4*) [88]. Within each subclass, these compounds may differ in the position and number of hydroxyl, methoxyl and glycosyl groups (*Fig. 4*). The latter ones may be further substituted (e.g. glycosylated or acylated), sometimes resulting in very complex structures. Some of the most common flavonoids are quercetin, a flavonol abundant in onion, tea, and apple; daidzein, the main isoflavone in soybean; hesperetin, a flavanone present in citrus fruits; cyanidin, an anthocyanin giving its color to many red fruits (blackcurrant, raspberry, strawberry, etc.) and catechin, a flavanol found in tea and several fruits [127].

Among flavanols, flavan-3-ols can be distinguished according to the hydroxylation pattern in ring A and B and the stereochemistry of the asymmetric C3 (*Fig. 3* and *4*) [130]. In nature, they exist either as monomers, oligomers or polymers (proanthocyanidins or condensed tannins). In certain positions, flavanol units may sometimes bear acyl or glycosyl substituents. The most common acyl substituent is gallic acid which forms an ester linkage with the hydroxyl group in the C3 position thus forming, by example, epigallocatechin and epigallocatechin gallate [89, 93].

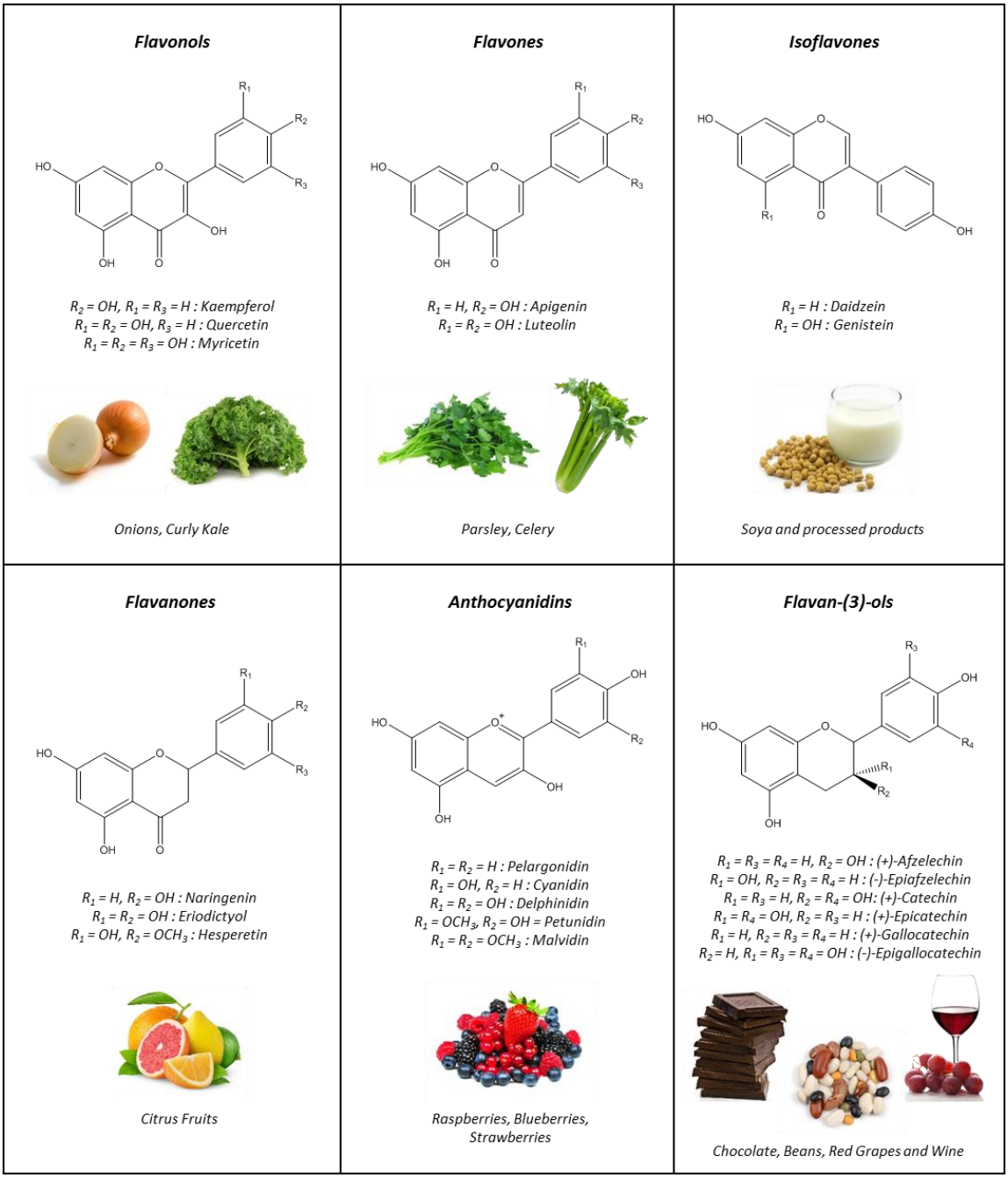


Fig. 4 – Chemical structures of the six major subclasses of flavonoids according to the type of heterocyclic C ring involved and its substitution pattern. Some examples of food products in which each class is abundant are also represented.

1.2.1 Tannins

The term tannin has been employed to designate substances of vegetable origin capable of transforming animal hides into leather [131]. They have been defined as water soluble phenolic compounds with molecular weights ranging from 500 to over

3000 Da that, besides giving the usual phenolic reactions, have special properties such as the ability to precipitate alkaloids, gelatin and other proteins [132, 133]. Nowadays, it is well established that soluble tannins extracted from plant tissues can reach molecular weights of several thousands (up to 30000 Da), depending on their chemical structure and also on their colloidal behavior in aqueous solution [89, 130, 134-136]. Furthermore, in addition to soluble tannins that can be readily extracted with different aqueous and organic solvents like aqueous methanol or acetone, insoluble forms that are resistant to all kinds of solubilization have also been identified. In that way, it has been that proposed tannin polymers become insoluble due to their high molecular size as well as their complexation with proteins or cell wall polysaccharides [89, 135, 137].

Concerning their chemical structure, tannins can be divided into two major groups: condensed tannins (proanthocyanidins) and hydrolysable tannins. In the diet, proanthocyanidins are more common than hydrolysable tannins [130].

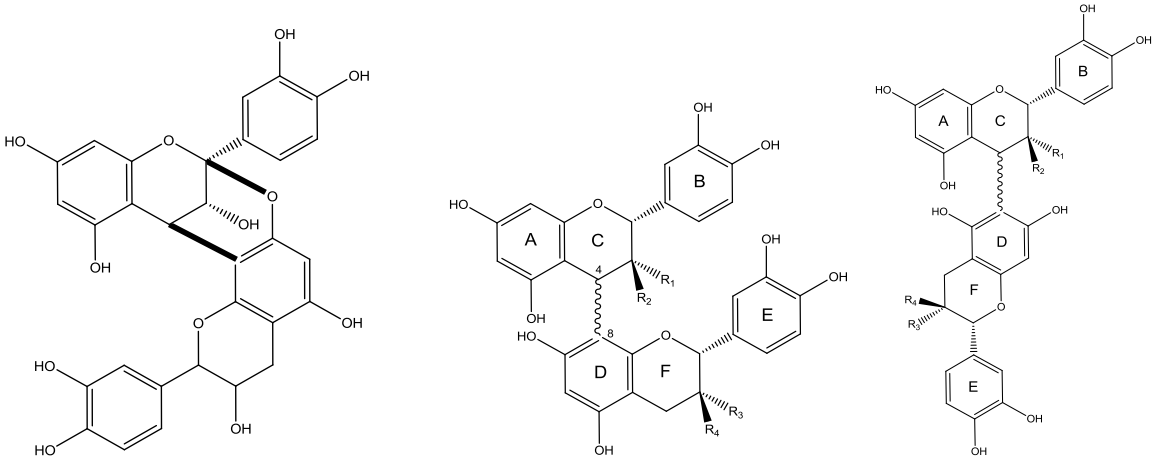
1.2.1.1 Condensed Tannins (Proanthocyanidins)

Condensed Tannins (also named proanthocyanidins) are polymers of flavan-3-ols whose elementary units are linked by C-C interflavanol bonds, established between the C4 of one flavan-3-ol unit and the C6 or C8 of another unit (B-type proanthocyanidins, *Fig. 5*) . In some plant compounds, an additional ether linkage between the C2 of the upper unit and the oxygen-bearing C7 or C5 of the lower unit also occur (A-type proanthocyanidins, *Fig. 5*) . Their degree of polymerization varies over a broad range of molecular forms from dimers up to about 200 monomeric flavan-3-ol units [138]. Depending on the B ring hydroxylation pattern of the chain-extender units, B-type proanthocyanidins, can be classified as propelargonidins (mono-hydroxylated), procyanidins (di-hydroxylated) or prodelphinidins (tri-hydroxylated) (*Fig. 6*) . In certain positions, proanthocyanidins may sometimes be esterified with gallic acid or exceptionally with sugars.

Because of the high diversity of tannin structures and molecular weights as well as the lack of suitable analytical methods to separate and quantify highly-polymerized tannins, data on the content and composition of these dietary compounds in foodstuffs is quite limited [130]. Therefore, the analysis of proanthocyanidins is mainly restricted to simpler structures such as monomers, dimers and some trimers, despite most of polyphenols are present in nature as polymers [139, 140]. In general, proanthocyanidins are primarily found in fruits, especially berries, cocoa and some beverages like wine, beer and tea. Vegetables are not an important source of these

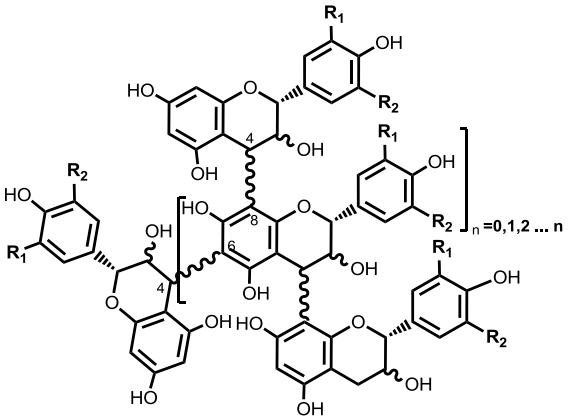
tannins. Legumes, nuts and other minority cereals such as sorghum and barley also contain proanthocyanidins, but they are not detectable in staple crops such as corn, rice and wheat. Proanthocyanidins (especially the most polymerized ones) tend to concentrate in the peel of fruits or the bran of grains [130, 141-143].

Procyanidin A₂ ***B-type Procyanidins***



R_1	R_2	R_3	R_4	C4-C8 DIMER NAME	C4-C6 DIMER NAME
OH	H	H	OH	*B1*	***B5***
OH	H	OH	H	*B2*	***B6***
H	OH	H	OH	***B3***	***B7***
H	**OH**	**OH**	**H**	***B4***	***B8***

Fig. 5 – Structures of simple dimeric procyanidins to demonstrate different linkage types in condensed tannins.



R_1 = H; R_2 = H : Propelargonidin
R_1 = OH; R_2 = H : Procyanidin
R_1 = OH; R_2 = OH : Prodelphinidin

Fig. 6 – General structure of proanthocyanidins.

1.2.3.2 Hydrolysable Tannins

Hydrolysable tannins are polyesters of a sugar moiety and organic acids. The term "hydrolysable tannin" comes from the fact that these compounds undergo hydrolytic cleavage to the respective sugar and acid moiety upon treatment with diluted acids. Usually, the sugar component is glucose, but fructose, xylose, saccharose among other structures is also found. If the acid component is gallic acid, these compounds are named gallotannins (*Fig. 7A*). Esters with hexahydroxydiphenic acid (forming ellagic acid upon hydrolysis) are called ellagitannins (*Fig. 7B*) [130, 144, 145]. Most ellagitannins are mixed esters both with gallic acid and hexahydroxydiphenic acid.

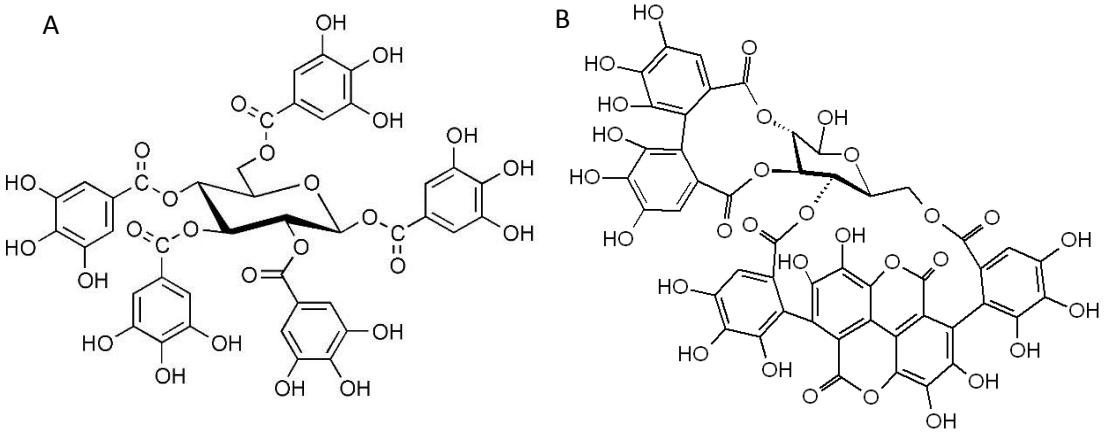


Fig. 7 – Examples of structures of hydrolysable tannins: pentagalloylglucose (gallotannin) (A), punicalagin (ellagitannin) (B).

Berries, legumes and leafy vegetables are the major sources of hydrolysable tannins. They are also found in peanuts, walnuts, pecans, cashews, pormegranate, red apples, kiwi, etc [130]. Even though hydrolysable tannins are essentially present in the non-edible portions of plants such as roots and branches, these compounds can be introduced in the diet by technological operations. In wines, for example, hydrolysable tannins can pass from the wood to wine during its aging in barrels. However, their intake from wine should be residual given the low amount of ellagic tannins on grapes and also because most of the wines are not matured in new oak wood in which there is a high content of ellagic tannins.

1.2.3 Interaction between Tannins and Proteins

The interaction between dietary polyphenols and proteins has been extensively studied from both a nutritional and pharmaceutical perspective, being very relevant for some sensorial and biological properties of those. Based on many prominent research works, as well as a result of the large improvement of analytical techniques, tannins are now described to establish cross-links with proteins by different kinds of bonds [89]:

- Hydrogen bonds between the hydroxyl groups of phenolic compounds and the carbonyl and -NH_2 functional groups of proteins [146-148];
- Van der Waals interactions supported by the hydrophobic effect between the benzene rings of phenolic compounds and the apolar amino acid side chains [149-151];
- Ionic bonds between the phenolate anions and cationic sites of proteins [152, 153];
- Covalent bonds resulting from the reaction between the nucleophilic groups of proteins such as -NH_2 and -SH and, on one side, quinone groups resulting from phenolic oxidation [153-155] or, on the other side, carbocations resulting from acid-catalysed condensed tannins depolymerisation [156].

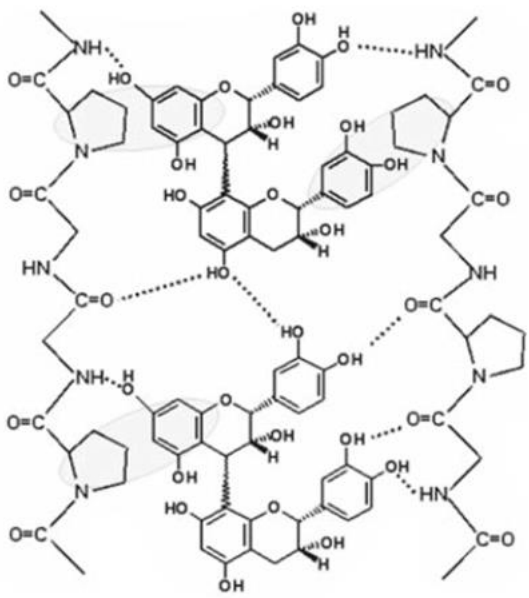


Fig. 8 – Scheme of the interaction between condensed tannins and proteins: main driving forces (hydrogen bonds and hydrophobic interactions) between phenolic rings (cross-linkers) of tannins and the amide groups and apolar side chains of amino acids such as proline. Adapted from Santos-Buelga *et al.* 2008.

Overall, hydrophobic interactions and hydrogen bonds have been reported as the main driving forces involved in tannin-protein binding (*Fig. 8*) [157]. The possibility of ionic interactions has been excluded at both acidic and neutral pH values because of the absence of charged groups at pH considerably below to the pKa values of the phenolic groups (~ 9-10) [158].

Using different yet complementary physico-chemical methods based on nuclear magnetic resonance (NMR), microcalorimetry, microscopy, high performance liquid chromatography (HPLC), fluorescence, mass spectrometry, infrared spectroscopy,

molecular modeling and light scattering experiments (turbidimetry, nephelometry and dynamic light scattering) to study the molecular mechanisms (structure-activity relationships) involved in the tannin-protein binding process, some reports have contributed to indicate that it may occur in a specific and selective manner [89, 147, 159]. In that way, several structural and environmental variables known to drive these interactions have been extensively assayed and are still being presently studied by some research groups. Concerning the influence of the protein, the interaction can be affected by its size [147, 160], charge [147], presence and type of side chains [161, 162] and conformation [163]. In general, it has been found that proteins which are readily precipitated by tannin are large, have a high proline content, and lack of secondary or tertiary structure, although some of them may possess a polyproline helix [164]. For complexation to occur, both tannin and protein must have the appropriate steric structure and molecular weight. Additionally, there are many works reported in the literature that have highlighted the strong influence of the polyphenol structure (hydroxylation degree, size, conformation, flexibility and others) on their interaction with proteins [147, 149, 165-168]. Overall, the affinity of tannins for proteins increases with the molecular weight (or polymerization degree) and the degree of galloylation apparently because the number of interaction sites increase with size. This behavior seems to be independent of the protein structure. Small changes in the structure of polyphenols such as the carbon configuration, the degree and pattern of hydroxylation were also shown to affect their affinity towards proteins [89].

The composition of tannin-protein aggregates depends not only on the nature of the species involved but also on their relative concentration (ratio) that will interfere in the network of the bonds established between them and, as a result, in the stoichiometry of the formed complexes. Classically, it was described that the ability of polyphenols to bind and precipitate proteins is favored by their aptitude to work as multidentate ligands that bind simultaneously to different sites of the same protein and also cross-link separate protein molecules [169, 170]. In that model, at low tannin concentration, each polyphenolic molecule associates to one protein to form small and soluble compacted complexes that are less hydrophilic than the protein itself. When increasing concentrations of polyphenol are added, tannins cross-link proteins leading to the formation of high soluble tannin-protein aggregates up to a maximum from which these become insoluble and precipitate (*Fig. 9*) [161]. Proteins can also wrap around tannins and these latter have the ability to auto-associate with other tannins, forming stacks, even when they are bound to proteins [149]. It has been proposed that the formation of the largest amount of aggregates occurs when the number of polyphenol binding ends and protein binding sites are nearly equal [171]. Protein precipitation by tannins may be

reversed by the addition of further tannin or protein [135, 169]. Nevertheless, PRPs are described as effective precipitants that are unable to re-solubilize the polyphenol molecules when present in excess [172, 173].

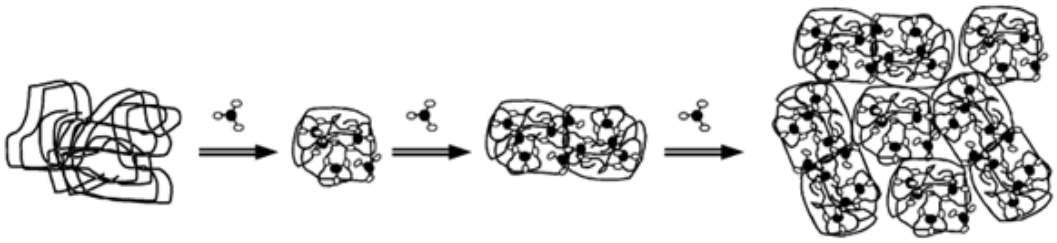


Fig. 9 – Proposed three-step model for the interaction between randomly coiled proteins and multidentate polyphenols. Adapted from Jobstl *et al.* 2004.

External factors such as pH, ionic strength, solvent composition (e.g. presence of ethanol that changes solvent polarity) and temperature also affect the complexation and precipitation of protein by vegetable tannins. The solvent pH is directly involved in the interaction between tannins and proteins since it affects the ionization degree of both elements. In general, the degree of binding and precipitation decreases at low pH (< 2) and at high pH (> 8) and is maximal close to the isoelectric point where electrostatic repulsions between proteins are minimized [89, 147].

1.2.4 Tannin Bioavailability

To exert any biological effect, dietary tannins have to be available to some extent in the target tissues. This feature critically depends on their absorption in the gut and their bioavailability. In fact, phenolic compounds that are usually common in the human diet are not necessarily the most active within the body, either because they have a lower intrinsic activity or because they are poorly absorbed from the gastrointestinal tract, highly metabolized, or rapidly eliminated [88, 130]. Using both human and animal models, the absorption, bioavailability and metabolism of monomeric phenols have been widely studied [174-177]. However, little is known about the bioavailability of polymeric tannins with the obtained results being controversial. Being quite unlikely that high-molecular-weight tannins are absorbed intact, it is supposed that the polymerization degree as well as the solubility of food tannins may have a major impact on their fate in the body. Therefore, highly polymerized tannins typically exhibit low bioaccessibility in the small intestine and low fermentability by colonic microflora [130, 178].

Several attempts have been made, by means of *in vitro* and *in vivo* studies, to determine the extent of hydrolysis of polymeric proanthocyanidins in the gastrointestinal tract as well as the possibility of oligomeric molecule absorption in the small intestine. While some *in vitro* experiments suggested that procyanidins from chocolate are hydrolyzed to simple and bioavaliable flavanol units in acidic conditions (similar to those of the human stomach) [179], other studies have supported the view that oligomeric proanthocyanidins are not significantly depolymerized into monomeric flavan-3-ols during their passage through the gastrointestinal tract [180, 181]. On the other hand, during small intestinal digestion, high-molecular-weight proanthocyanidins can aggregate with several dietary proteins, starch and digestive enzymes, leading to the formation of complexes that are less soluble and less accessible to enzymes [130, 182]. Therefore, within those supramolecular structures, *in vitro* studies revealed that digestive enzymes were not able to release and increase the bioaccessibility of proanthocyanidins from the food matrix, suggesting that highly polymerized proanthocyanidins may reach the colon unchanged [130, 143]. Small-molecule proanthocyanidins such as dimmers and trimers present less complex formation activity allowing them to be readily absorbed [183].

Another important site where dietary tannins become accessible in the gastrointestinal tract is the large intestine. In there, the abundant bacterial microflora displays a major role in their metabolism [178, 184]. After becoming fermentable substracts for microbial enzyme metabolism, there are two possible routes available for any tannin that reaches the colon: breakdown of the original tannin structure into absorbable metabolites or breakdown into non-absorbable metabolites that remain in the colonic lumen where they may counteract the activity of dietary pro-oxidants [126, 130]. Concerning (polymeric) proanthocyanidins, several authors have found that those are highly metabolized by the gut microflora. In fact, many metabolites were detected in rat's urine after consumption of proanthocyanidin dimers, trimer and polymers. These include aromatic acids such as phenylvaleric, phenylacetic, phenylpropionic and benzoic acid derivatives, the total yield of which decrease significantly according to the polymerization degree of the precursor proanthocyanidin. As only a small percentage of the microbial metabolites are excreted in urine, it is possible for proanthocyanidins to pass through the entire gastrointestinal tract largely intact, especially for those with a high molecular weight [130, 143, 178, 185]. Although bacteria which are resistant to proanthocyanidins antibacterial properties have been isolated from the gastrointestinal tract ecosystem, there are no reports of intestinal bacteria that can degrade proanthocyanidins and their monomeric flavan-3-ols units [130, 186].

Focusing on hydrolysable tannin bioavailability, few studies are available in which their hydrolysis rate into ellagic acid and/or gallic acid monomers was evaluated in a context of gastric/intestinal enzymatic digestion [187]. Nevertheless, the absorption of gallic and ellagic acid has been extensively studied. In the case of hydrolysable tannins, their digestion involves the breakdown of the bonds between the phenolic molecules and sugars with the production of free sugars and simple phenolic acids that can be absorbed in the duodenum or metabolized by the microbial flora in the colon [188].

1.2.5 Tannin Biological Effects

Nowadays, several healthy effects have been attributed to the intake of tannins, especially proanthocyanidins. Despite the lack of precise knowledge about the fate of these compounds in the human body, by now, it is certain that tannins are partially metabolized and available for absorption at different sites of the gastrointestinal tract. Therefore, it is believed that tannins may exert their biological effects in two different ways: as a non-absorbable and complex structure with binding properties that might produce local effects in the gastrointestinal tract (antioxidant, radical scavenging, antimicrobial, antiviral, antimutagenic and possibly the modulation of cellular signaling pathways) or as absorbable tannins (those with a low molecular weight) and metabolites from colonic fermentation of tannins which should produce systemic effects in numerous organs and tissues. A different, chemical approach was proposed by some authors to explain the biological properties of dietary tannins. In that way, three distinct mechanisms have been suggested regarding tannin action mode: by complexation with metal ions, through antioxidant and radical scavenging activities or through their ability to complex with other macromolecules such as proteins and polysaccharides [93, 127, 130, 189].

EXPERIMENTAL

2.1 REAGENTS

All organic solvents used in this study were of analytical grade. Acetonitrile was purchased from Panreac while acetic acid (99.8%) was obtained from Sharlau. Trifluoroacetic acid and hydrochloridric acid (37%) were both obtained from Sigma-Aldrich. All reagents used in SDS-PAGE including trizma base (≥ 99.9%), glycine (99%), sodium dodecyl sulfate (99%), glycerol (99%), 2-mercaptoethanol (≥ 98%), bromophenol blue, imperial protein stain, acrylamide (≥ 99%), N,N'-methylenebisacrylamide (98%), N,N,N',N'-Tetramethyl-ethylebediamine and ammonium persulfate (98%) were also from Sigma-Aldrich. Sinapinic acid (SA), 2,5-dihydroxybenzoic acid (DHB) and α-cyano-4-hidroxycinnamic acid (CHCA) from Sigma-Aldrich were the matrices used in MALDI-TOF-MS analysis. Pepsin from porcine gastric mucosa (CAS 9001-75-6), pancreatin from porcine pancreas (CAS 8049-47-6), α-chymotrypsin from bovine pancreas (CAS 9004-07-3) and gliadin from wheat (CAS 9007-90-3) were also acquired from Sigma-Aldrich.

The peptide QLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF (33-mer peptide without the initial leucine residue), henceforth CRP32, was synthetized and purchased from ChinaPeptides CO. (Pudong New Area, Shanghai, China) and had a purity > 95%, as assessed by HPLC-ESI-MS.

2.2 GRAPE SEED TANNIN ISOLATION

Condensed tannins were extracted from *Vitis vinifera* grape seeds, according to the method described in the literature [190, 191]. After a liquid-liquid extraction with an ethanol/water/chloroform solution (1:1:2), the resulting mixture was centrifuged and the chloroform phase, containing chlorophylls, lipids and other undesirable compounds, was rejected. The hydroalcoholic phase was then extracted with ethyl acetate and evaporated using a rotary evaporator (35 °C), yielding a residue composed of monomeric and oligomeric procyanidins. These compounds were fractionated though a TSK Toyopearl HW-40(S) gel column (100 mm x 10 mm i.d., with 0.8 mL/min of methanol as eluent) producing five fractions. Fractions I to IV were obtained after elution with 99.8% (v/v) methanol for 30 min, 45 min, 1h and 5 h respectively while fraction V was obtained after elution with methanol/ 5% acetic acid (v/v) during the next 14 h. All fractions were mixed with deionized water, the organic solvent eliminated using a rotary evaporator under reduced pressure (35 °C)

and then freeze-dried. The procyanidin composition of each fraction was determined by direct analysis through Electrospray Ionization Mass Spectrometry (ESI-MS) (Finnigan DECA XP PLUS) [192].

2.3 PROCYANIDIN B3, PROCYANIDIN TRIMER T1 AND PROCYANIDIN TETRAMER TT1 SYNTHESIS

Procyanidin B3, procyanidin trimer T1 and procyanidin tetramer TT1 were obtained by hemisynthesis using (+)-taxifolin and (+)-catechin [193, 194]. Briefly, taxifolin and catechin (ratio 2:1) were dissolved in ethanol under argon atmosphere. The mixture was then treated by dropwise addition of sodium borohydride, and left for 15 min under magnetic agitation. The pH was lowered to 4.5 by slowly adding acetic acid/water 50% (v/v) and the mixture was allowed to stand under argon atmosphere for 30 min. After this, the reaction mixture was extracted with ethyl acetate, evaporated and passed through C18 gel, thoroughly washed with water and recovered with methanol. The obtained fraction, after evaporation of methanol, was passed through a TSK Toyopearl HW-40(S) gel column (300 mm x 10 mm i.d., with 0.8 mL/min of methanol as eluent) coupled to a UV-Vis detector (Gilson 115) were several fractions were recovered and analyzed by ESI-MS (Finnigan DECA XP PLUS) yielding procyanidins with varying degrees of polymerization. The fractions containing procyanidin B3 ($[M-H]^-$ = 577), procyanidin trimer T1 ($[M-H]^-$ = 865) and procyanidin tetramer TT1 ($[M-H]^-$ = 1153) were isolated and freeze-dried. The purity of those fractions was assessed by LC-MS and direct MS analysis, and was higher than 95%.

2.4 CHARACTERIZATION OF WHEAT GLIADINS RAW EXTRACT

2.4.1 MALDI-TOF-MS Analysis. MALDI-TOF-MS was used to obtain the mass spectra of the commercial gliadin raw extract. For this, two microliters of raw sample (0.625 mg/mL), prepared in 62.5% ethanol/ 3.75% acetonitrile/ 0.0125% trifluoroacetic acid (v/v) aqueous solution, was mixed with 2 µL of matrix solution containing 10 mg/mL sinapinic acid in 30% acetonitrile/ 0.1% trifluoroacetic acid (v/v). Afterward, 2 µL of this matrix-sample mixture was applied onto a stainless steel target plate (MTP 394

target plate ground steel BC, Bruker Daltonics) and finally air-dried. The respective mass spectrum was recorded on an Ultraflextreme mass spectrometer (Bruker Daltonics) operating in linear positive ion detection mode with laser SmartBeam-III and under FlexCompass 1.4 software control (Bruker Daltonics). The mass spectrum was acquired in a range from 10 to 55 kDa (or 210 kDa), by accumulation of 500 laser shots with 1000 Hz of frequency. For analysis optimization, different analyte/matrix ratios were assayed (1:1, 1:2, 2:1). 2,5-Dihydroxybenzoic acid (DHB) was also tested as MALDI-TOF-MS matrix according to Bruker Daltonics preparation protocols.

2.4.2. Protein Separation by SDS-PAGE. The protein composition of the wheat gliadins raw extract was investigated using an automated capillary gel electrophoresis system (Power Supply 300V, from VWR). The method is based on a separation of proteins according to their sizes by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using a 10% acrylamide resolving gel and 5% acrylamide stacking gel. All samples (1.5 mg/mL) were prepared in 2x SDS protein sample buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 10% v/v β-mercaptoethanol and 0.04% bromophenol blue) and heated at 90°C for 5 min with shaking. BSA was used as an internal control. The running buffer was 250 mM Tris-HCl pH 8.3, 1.92 M Glycine and 1% SDS. Molecular weight markers were broad ranged (Precision Plus ProteinTM Unstained Standards, Bio-Rad). The separation was achieved by an electric current of 140V. After electrophoresis, the gels were stained with Imperial Protein Stain - a Coomassie R-250 dye-based reagent - for 2 h. The destaining step was done by washing the gels overnight with water:methanol (50:50 v/v).

2.4.3. Tryptic digestion. After eletrophoresis, the bands of interest were excised from the gel and transferred to a rack. The gel pieces were washed twice with 25 mM ammonium bicarbonate/ 50% acetonitrile and one time with 100% acetonitrile. Once washed, the pieces were dried into a SpeedVac (Thermo Savant). Forty microliters of 10 µg/mL trypsin in 50 mM ammonium bicarbonate was added to the dried gel and the samples were incubated overnight at 37ºC. After the incubation, the extraction of tryptic peptides was performed by the addition of 10% formic acid/ 50% acetonitrile (three times) followed by liophilization. Tryptic peptides were resuspended in 50% acetonitrile/ 0.1% formic acid/ 49.9% water and subsequently analyzed by mass spectrometry (ESI-MS/MS, Nano-ESI-MS/MS and MALDI-TOF-MS).

2.4.4. Protein Identification by Mass Spectrometry.

2.4.4.1. ESI-MS/MS Analysis. 15 µL of each tryptic digest (three replicates per sample) were injected into a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) after their chromatographic separation into an Accela HPLC system (Thermo Fisher Scientific). The chromatographic conditions are described in *Table 2* and the ESI-MS/MS conditions are displayed in *Table 3*.

Table 2 – Chromatographic conditions for the identification of gluten proteins.

CHROMATOGRAPHIC CONDITIONS	
INJECTION	15 µL
GUARD COLUMN	Pelliguard LC-8 (5 mm × 4.6 mm i.d.<comma> 40 µm) (Supelco)
ANALYTICAL COLUMN	Water symmetry C18 (150 mm × 4.6 mm i.d.<comma> 5 µm) (Waters).
MOBILE PHASE	A: 0.8% formic acid in water (v/v); B: 1% formic acid in acetonitrile (v/v) 0–2 min 15% B, 40 min (elution step) 15–30% B, 40-42 min (elution step) 30% B, 50-52 min (washing step) 100%B, 52-60 min (conditioning step) 15% B
FLOW	500 µL/min
DETECTION CONDITIONS	
SCANNING	200–300 nm
DETECTION WAVELENGTH	214 nm
SCAN RATE	20 Hz
STEP	1 nm
BANDWIDTH	9 nm

Table 3 – ESI-MS/MS conditions for the identification of gluten proteins.

ELECTROSPRAY CONDITIONS	
CAPILLARY TEMPERATURE	275 °C
CAPILLARY VOLTAGE	35 V
SHEATH GAS FLOW	5 arbitrary units
SPRAY VOLTAGE	3 kV
TUBE LENS	100 V
MS/MS CONDITIONS	
MODE	CID
NCE	35 %
The MS^2 was applied for the three most intense peaks	

2.4.4.2. Nano-ESI-MS/MS analysis. Given the lack of sensitivity of the ESI-MS/MS approach, Nano-ESI-MS/MS conditions were tried. The samples were pumped into a UHPLC Dionex UltimateTM 3000 RSLCnano LC system (Thermo Scientific) with EASY-nano spray following the conditions showed in *Table 4*.

Table 4 - Nano-ESI-MS/MS conditions for the identification of gluten proteins.

CHROMATOGRAPHIC CONDITIONS	
INJECTION	20 µL
COLUMN	EASY-spray PepMap C18 (15 cm x 75 µm ID) 3 µm 100A (Thermo Scientific)
MOBILE PHASE	A: 0.1% formic acid in water (v/v); B: 80% acetonitrile/ 19.92% water/ 0.08% Formic acid (v/v) 0–3 min 5% B, 60 min (elution step) 55% B, 65 min (elution step) 90% B, 65-75 min (washing step) 90% B, 80 min (conditioning step) 5% B
FLOW	0.3 µL/min
ELECTROSPRAY CONDITIONS	
CAPILLARY TEMPERATURE	250 °C
CAPILLARY VOLTAGE	9 V
SHEATH GAS FLOW	5 arbitrary units
SPRAY VOLTAGE	1.9 kV
TUBE LENS	100 V
MS/MS CONDITIONS	
MODE	CID
NCE	35 %

When the run was completed, Proteome Discoverer 1.4 software (Thermo Fisher Scientific) was used to interrogate protein databases for protein(s) identification. SEQUEST search algorithm was used after limiting the search to the wheat gluten proteome (UniProtKB database). The resulting search outputs were evaluated in terms of the corresponding score values and protein coverage.

2.4.4.3. MALDI-TOF-MS Analysis. In order to confirm the proteins identity, MALDI-TOF-MS mass spectrometry was tested as an alternative method. Thus, 1 µL of α-cyano-4-hydroxycinnamic acid (CHCA) at 0.7 mg/mL was mixed with 1 µL of each tryptic digested sample (2 µg/mL), both prepared in 30% acetonitrile/ 0.1% TFA aqueous solution. This was performed according to Bruker Daltonics preparation protocols. These mixtures were then layered over three replicate spots on the stainless steel target plate, air-dried, and further analyzed using a Bruker Daltonics MALDI-TOF-MS/TOF instrument, operating in positive-ion reflector mode with laser SmartBeam-III and under FlexCompass 1.4 software control (BrukerDaltonics). For each triplicated sample, the mass spectrum was automatically acquired from 500 to 5000 Da in which 1000 laser shots with 2000 Hz of frequency were accumulated.

2.5 SEPARATION AND IDENTIFICATION OF GLIADIN DERIVED PEPTIDES

2.5.1 In vitro Digestion of Gliadin Raw Extract. The commercial raw extract of wheat gliadins was subjected to enzymatic digestion, according to the method described in the literature [10]. Briefly, to simulate a gastric digestion, 100 mg of gliadins were weighed and dissolved in 0.01 M hydrochloridric acid followed by incubation in a 37 °C water bath with pepsin (1:100 protease to protein w/w ratio) at pH 2.0 for 30 min. The reaction mixture was then pH adjusted to 7.0 in 50 mM phosphate buffer and added with pancreatin (1:100 w/w) and chymotrypsin (1:100 w/w) at 37 °C for 24 h. After this, the resulting peptide mixture was divided in 2 mL aliquots and centrifuged at 10.5 rpm for 5 min on an Eppendorf Minispin centrifuge.

2.5.2 Fractionation of Peptides Obtained after Wheat Gliadin Digestion. The previously collected supernatants, containing a vast amount of peptides derived from the enzymatic digestion of wheat gliadins, were fractionated through semi-preparative HPLC using a C18 reversed-phase analytical column (Merck Lichrospher C18 ODS, 5 µm, 250 x 25 mm) on a Knauer K-1001equipment, with a UV-Vis L-2420 Merck Hitachi (Elite) detector. A linear elution gradient was implemented using two mobile phases: the aqueous A solvent consisted in 0.1% formic acid/ 0.025% trifluoroacetic acid (v/v) and the solvent B contained 80% acetonitrile in water. Absorbance was monitored at 280 nm, and the flow rate was maintained at 0.5 mL/min. The gradient applied was linear from 0 to 95% B for 40 min. After each run, the column was washed with 100% B for 5 min, and equilibrated with the starting B concentration for 10 min. Each one of the seven peptide fractions (*Pep Mix1* to *Pep Mix7*, *Fig. 17*), collected at different retention times, were then freeze-dried.

2.5.3 Peptide Sequence Determination by Mass Spectrometry. LC-MS/MS experiments were performed on a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) coupled with an Accela HPLC system (Thermo Fisher Scientific) equipped with a Merck Lichrospher C18 ODS column (5 µm, 250 x 25 mm). The MS instrument was set as aforementioned (2.4.4.1). All MS and MS/MS spectra were acquired in the data-dependent mode. The instrument executed one MS scan followed by an MS/MS scan of each one of the three most intense peaks. The mobile phases for LC

separation were (A) 0.1% (v/v) formic acid in water and (B) 0.1% (v/v) formic acid in acetonitrile. Those were pumped at 0.5 mL/min in a linear gradient from 15 to 30% B over 40 min. For protein identification, as the same for the gliadin raw extract protein identification, the LC-MS/MS data were used to search the Uniprot protein sequence database by using the Proteome Discoverer 1.4 search engine (Thermo Fisher Scientific). FASTA files corresponding to wheat gluten proteome were taking into consideration for the identification process.

2.6 PEPTIDE-TANNIN INTERACTION ASSAYS

2.6.1 Fluorescence Quenching Measurements. The quenching effect between two selected peptide mixtures (*Pep Mix4* and *Pep Mix6*) and different polyphenols (procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and fraction V of oligomeric procyanidins) was assayed using a Perkin-Elmer LS 45 fluorimeter. For the fluorescence quenching measurements, tryptophan was used as an intrinsic fluorophore. The excitation wavelength was set to 290 nm and the emission spectrum was recorded from 300 to 500 nm. Both slits were 10 nm. All experiments were performed in Milli-Q ultrapure water. In several 2 mL microtubes, increasing volumes of different polyphenols stock solutions (100 µM) were added to the peptide mixtures assay solutions (0.2 mg/mL), in order to give final concentrations of polyphenols in the range of 0 to 25 µM. After 30 min of reaction, the microtubes were shaken and the emission spectra were measured in the fluorimeter cell. Between each experiment, the cell was washed three times with ethanol and water. Since procyanidins absorb energy at the established emission wavelength [195] a blank was made for each polyphenol concentration, in which the peptide solution was replaced by Milli-Q ultrapure water. The respective spectra were then automatically subtracted from the emission spectrum of the corresponding solution [167, 196]. The possibility of fluorescence resonance energy transfer (FRET) between the peptide mixtures and the tested polyphenols was discarded after analysis of both absorption and emission spectra.

The fluorescence lifetimes of *Pep Mix4* and *Pep Mix6* were measured on a Fluoromax-4 spectrophotometer, attached to a single photon counting controller (FluoroHub), both from Horiba Jobin-Yvon, at room temperature. The fluorescence excitation was performed with a Horiba Nano LED source of 290 nm, and fluorescence emission was recorded at the maximum wavelength for each peptide mixture (360 nm

for both solutions). The lamp profile was recorded by placing a scatter (dilute solution of LUDOX in water) in place of the sample [197].

2.6.2 Dynamic Light Scattering Measurements. The size of the peptide-tannin aggregates in solution was determined by dynamic light scattering (DLS) using a Zetasizer Nano ZS (Malvern). In this device, the sample solution was illuminated by a 633 nm laser, and the intensity of light scattered at an angle of 173° was measured by an avalanche photodiode. Differences in the aggregation process depending on tannin or protein concentration were evaluated. Hence, two different peptide concentrations (0.02 and 0.2 mg/mL) as well as different tannin concentration ranges were tested. Moreover, various tannins were assayed: procyanidin B3, procyanidin tetramer TT1 and an oligomeric mixture of tannin procyanidins (FV).

Different volumes of tannin stock solutions were mixed with a fixed volume of peptide fraction stock solution (*Pep Mix1* to *Pep Mix7*) and allowed to react for 30 min. After this, the mixture was shaken, transferred to a DLS plastic disposable cell, and the measurement was performed. All solutions were filtered through 0.2 µm disposable PTFE filters before mixing. Each experiment was performed in Milli-Q ultrapure water.

2.6.3 Identification of Peptide-B3 Complexes by Mass Spectrometry. The complexes formed between procyanidin B3 and gliadin-derived peptides as well as with CRP32 were characterized by ESI-MS. The method used for identification of peptide-tannin complexes was adapted from previous studies [120, 121]. In several microtubes, an aliquot of procyanidin B3 (B3) at 1 mM was added to a fixed volume of different peptide fractions or CRP32 (1 mg/mL), in order to achieve a final concentration of 500 µM of B3 and 0.2 mg/mL of peptide(s). Prior to analysis, the samples were vortexed for 2 min and maintained at room temperature for 30 min. All experiments were performed in Milli-Q ultrapure water with 0.1% of formic acid. After this, each B3-peptide(s) mixture was directly pumped at 5 µL/min into the LTQ-Orbitrap XL, controlled by Xcalibur 2.2 software. The ESI conditions were standardized as 3 kV of spray voltage, 35 V of capillary voltage and a sheath gas (N_2) flow of 5 a.u (equipment arbitrary units). The temperature in electrospray chamber was maintained at 270 ºC and the tube lens inside the mass analyzer was performed as 110 V.

For each mixture, the structure of the newly formed B3-peptide complexes was hypothesized by matching the new mass values with the different peptides exact mass in which successive increments of procyanidin B3 units added a fixed m/z value.

Common adducts (mainly Na^+ and K^+) and loss of ions such as H_3O^+ or NH_4^+ have been considered.

2.7 STATISTICAL ANALYSIS

All assays were performed at least in n = 3 repetitions. Values are expressed as the arithmetic means ± SD. Statistical significance of the difference between various groups was evaluated by one-way analysis variance (ANOVA) followed by the Tuckey test. Differences were considered to be significant when $P < 0.05$. All statistical data were processed using the GraphPad Prism 6.1 (GraphPad Software, San Diego, USA).

Results and Discussion

3.1 GRAPE SEED TANNIN FRACTIONS IDENTIFICATION

Oligomeric procyanidins from *Vitis vinifera* grape seeds (*Fig. 10*) were extracted and fractionated through column chromatography, yielding five tannin fractions comprising procyanidins with varied polymerization degrees. The composition of each recovered fraction was determined by direct analysis through ESI-MS and the mean molecular weight of those was estimated based on the relative abundance of each flavanol present (*Table 5*).

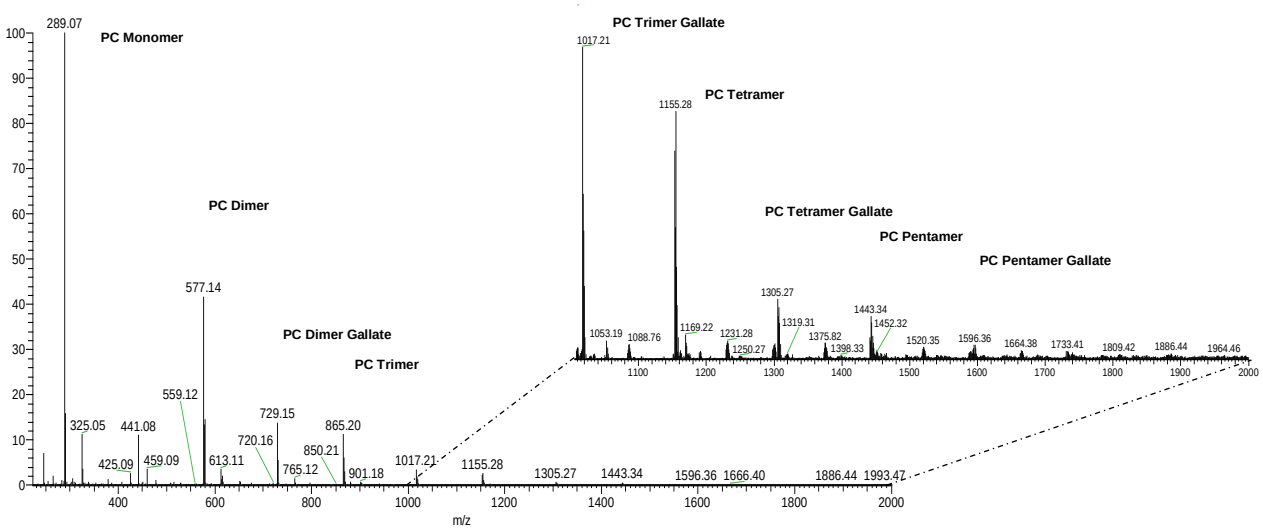


Fig. 10 – ESI-MS spectra of oligomeric procyanidins from grape seeds. To highlight and differentiate the high molecular weight procyanidins, the original spectra was enlarged between 1000 and 2000 m/z.

Table 5 – ESI-MS analysis of the five oligomeric procyanidin fractions. The mean molecular weight of the fractions was determined based in the relative abundance of each flavanol present.

	MOLECULAR WEIGHT RANGE (m/z) $[M-H]^-$	MEAN MOLECULAR WEIGHT
PC FRACTION I	169 - 729	290
PC FRACTION II	289 - 729	457
PC FRACTION III	729 - 1153	862
PC FRACTION IV	865 - 1881	1271
PC FRACTION V	1017 - 2169	1524

As accessed, fractions I to IV contained essentially catechin monomers and procyanidin oligomers up to pentameric molecules while fraction V contained mainly

mono- and digalloylated procyanidin pentamers, hexamers and galloylated procyanidin heptamers. According to *Table 5* and solvents used, it was found that the procyanidins molecular weight increased with their elution order with the last collected fractions becoming successively enriched in more complex and oligomerized tannin molecules. In the current study, only fraction V was used in the subsequent fluorescence quenching and dynamic light scattering interaction assays.

3.2 CHARACTERIZATION OF WHEAT GLIADINS RAW EXTRACT

3.2.1 MALDI-TOF-MS Analysis of Wheat Gliadins Raw Extract. An ideal sample preparation for MALDI-TOF-MS analysis would be of a homogenous layer of small matrix crystals containing a solid solution of the analyte. Native gluten proteins, especially the high-molecular ones, cannot be easily solubilized. Therefore, their characterization in solution proves to be extremely difficult. Given the high complexity of the wheat proteome along with the extraordinary heterogeneity of its proteins, gluten analysis is dependent on a previous protein fractionation procedure. This separation is usually made according to the proteins solubility: the albumins and globulins are extracted using water and/or salt solutions while gliadins are recovered with aqueous alcohol (60-70%). Glutenins, on the other hand, should be extracted with an appropriated extraction buffer (50% 1-propanol, 2 M Urea, 0.05 M Tris-HCl (pH 7.5), 2% (w/v) dithiothreitol).

In a first step, this study focused on the characterization of the wheat gliadins raw extract so all of its components had to be considered for MALDI-TOF-MS analysis. Unless a good protein solubilization is achieved, then regardless on the technique used, the analysis of gluten will be deceptive. Therefore, obtaining a good solubility was the main priority. Detergents, such as sodium dodecyl sulfate (SDS), has been used to increase protein solubility [198]. Nevertheless, when using chaotropic detergents, the protein structure changes due to micelle formation. Up to 95% of flour proteins can be extracted by SDS [199]; however, the ability of SDS to form micelles with proteins disturbs the analysis of their size and the results obtained in the presence of this detergent are not reproducible [200] and could interfere in the MALDI-TOF-MS analysis. The use of acids, like acetic acid, has shown to be quite efficient in solubilizing gluten proteins. Indeed, up to 86% of gluten proteins can be solubilized in 0.01 M acetic acid [201]. Also, 1-propanol turned out to be very efficient in solubilizing

both gluten components: monomeric gliadins and polymeric glutenins [202]. Sonication is one of the methods that can be used to improve the solubility of large molecules. However, this method may induce the breakdown of large polymers, thus leading to an underestimation of their sizes. Changes in protein conformation can also ocur upon sonication, so it had to be discarded. All the aforementioned solvents were tried to completely dissolve the raw extract constituents. Surprisingly, the only solvent mixture that leaded to a complete protein solubilization was achieved with the ones used for preparation of the MALDI-TOF-MS matrices.

In order to achieve best results, after samples solubilization, there is a choice of different matrices as well as preparation techniques that may be applied, both of which depends on the nature of the analyte to be detected. Here, aim was to obtain a homogenous preparation of the matrix, both in terms of sample distribution and geometry. Two different matrices were tested: sinapinic acid (SA) and 2,5-dihydroxybenzoic acid (DHB). In general, SA is a useful matrix for the analysis of peptides (> 3 kDa), high molecular-weight proteins (10-150 kDa) and some polar polymers. However, SA tends to form adducts with the analyte ions. These adducts can be resolved in the mass spectrum of proteins up to 40 kD. 2,5-DHB, on the other hand, is commonly used for MALDI-TOF-MS analysis of a wide variety of peptides, phosphopeptides, proteins, glycoproteins, polymers, carbohydrates, lipids and glycolipids.

MALDI-TOF-MS-MS analysis can be divided into three components: binding of wheat prolamins to the matrix, ionization and desorption of the prolamins by a laser separation (nitrogen gas UV laser with a wavelength of 337 nm, repetition rate of 1–20 Hz and pulse energy of 120–200 μJ), and its detection by a mass spectrometer. In order to develop an optimum MALDI-TOF-MS method, different parameters, namely sensitivity and resolution, were evaluated under the different sample-preparation conditions. Using the wheat gliadins raw extract, similar protein profiles were obtained within both matrices. (*Fig. 11*).

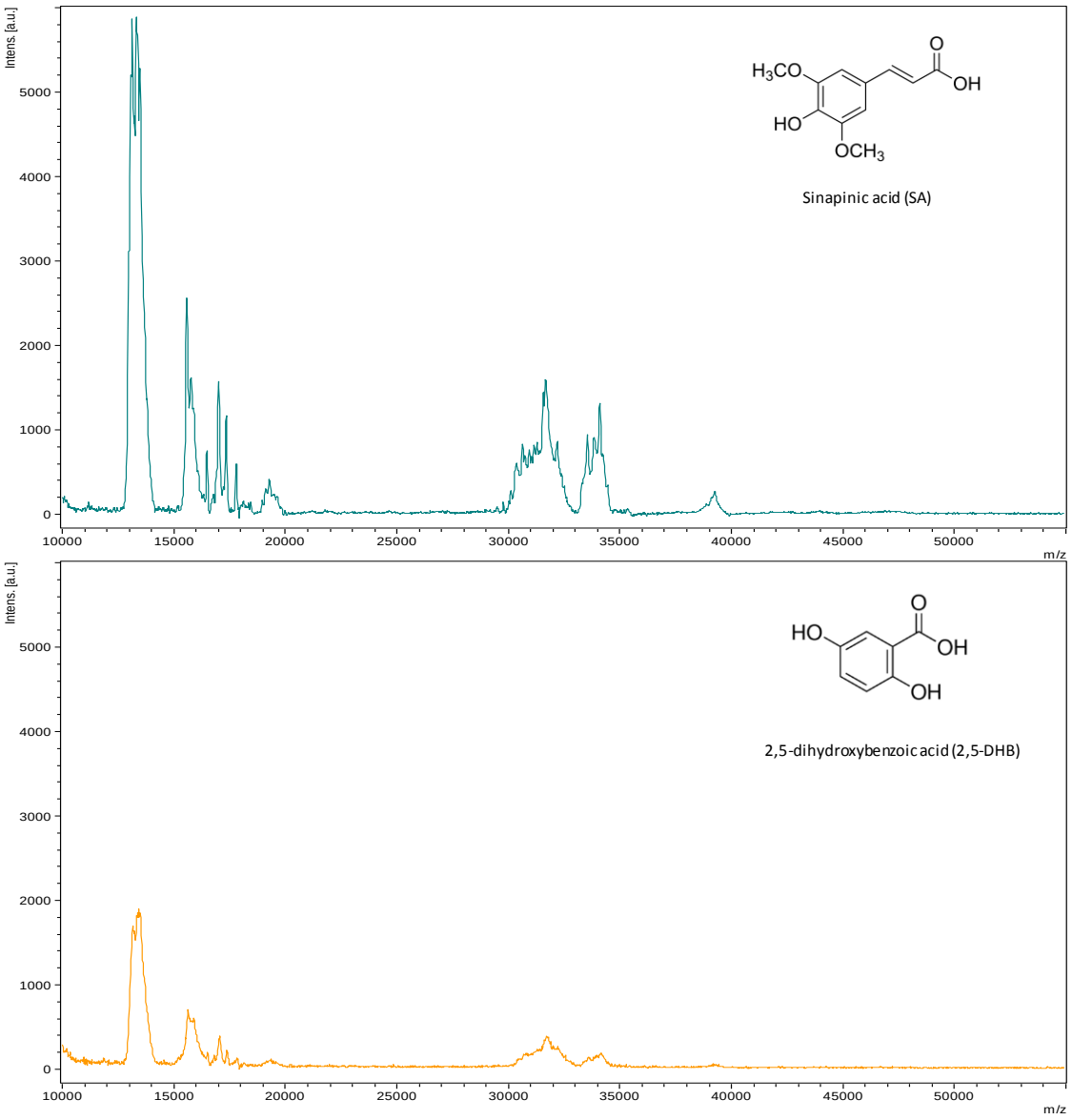


Fig. 11 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract (10-55 kDa) using both sinapinic acid (SA) and 2,5-dihydroxybenzoic acid (2,5-DHB) as matrices.

The main differences were found between resolution and sensitivity for low and high masses depending on the matrix used. Moreover, the main disadvantage of DHB was the fact that it forms big crystal needles. This feature contributed to a lack of sample homogeneity.

According to *Fig. 11*, it was the SA that produced the best results in the whole range of analysed mass values (10-55 kDa). The obtained signal was significantly higher when SA was tested (from twice to ten-fold). Furthermore, in SA assays, the resolution values were around 10.000 Da whereas by using DHB, not even half of this value was achieved. Additionally, different analyte/matrix ratios were tested (1:1, 1:2, 2:1) resulting in the 1:1 and 1:2 ratios as the best choices (*Fig. 12*).

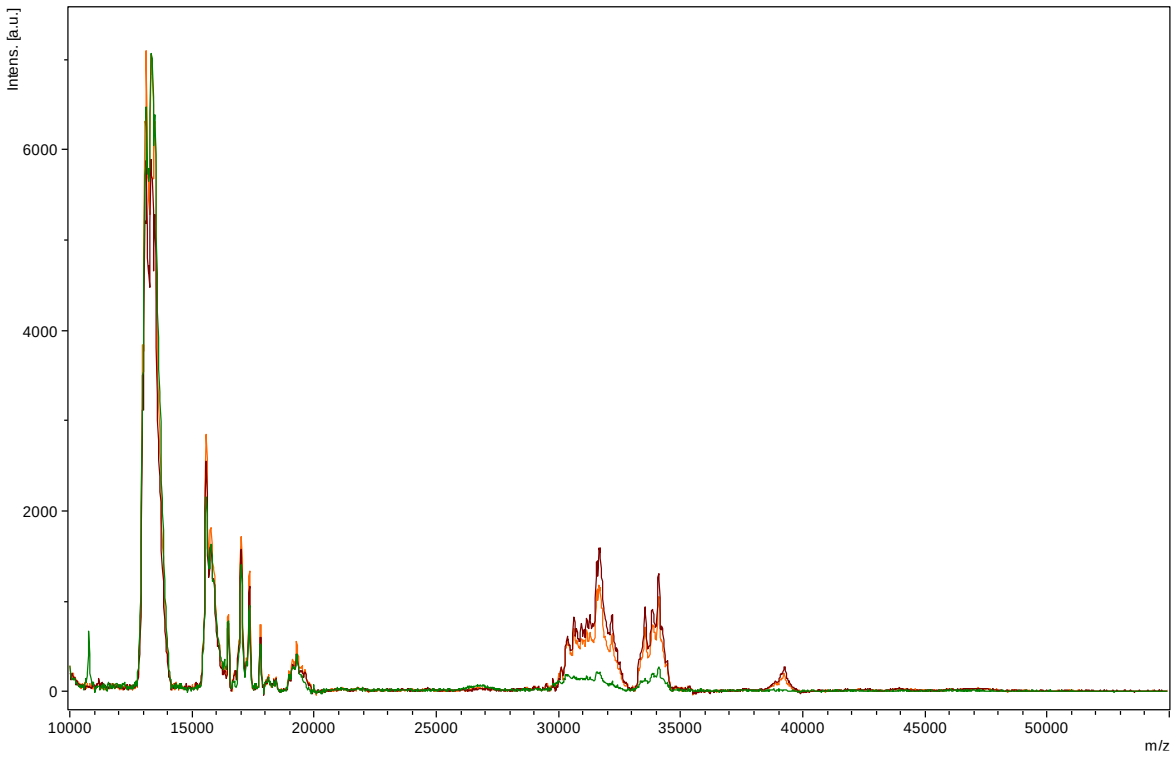


Fig. 12 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract using different analyte/SA ratios: 1:1 (orange color), 1:2 (brown color) and 2:1 (green color).

The mass spectrum of the wheat gliadins raw extract, after adjusting MALDI-TOF-MS experimental parameters, is shown in *Fig. 13*. Twenty-two peaks have been detected from 10 to 20 kDa, whereas eighteen peaks were noticed in a range of 20 to 30 kDa. Unresolved molecular entities above 60 kDa were also perceived in raw sample, when the mass range was amplified to 140 kDa (*data not shown*).

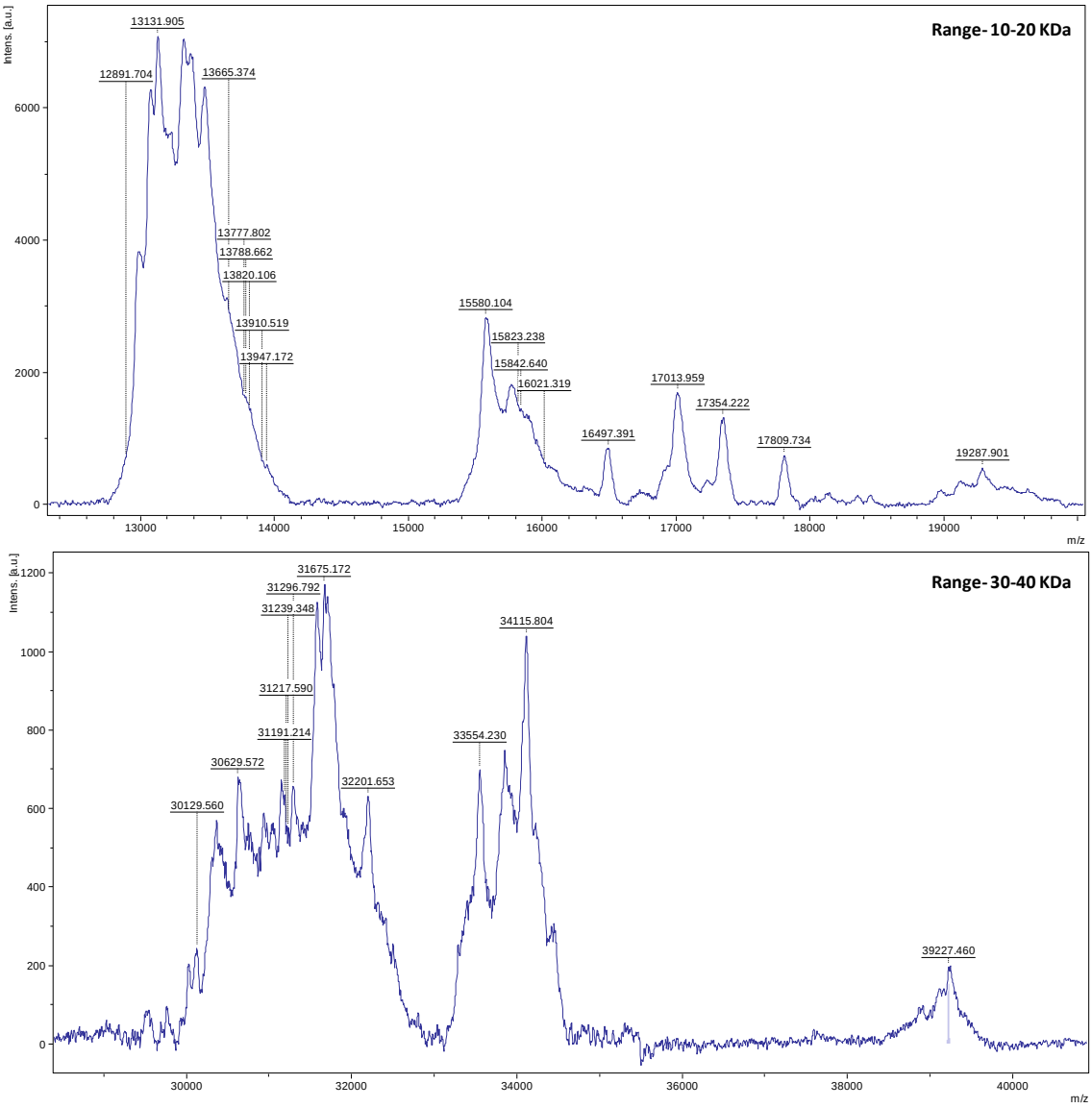


Fig. 13 – MALDI-TOF-MS mass spectra of the wheat gliadins raw extract

According to previous studies, despite the great heterogeneity and high genetic variability that defines the different classes of wheat gliadin proteins, the obtained mass spectrum clearly shows their protonated mass pattern around 29-42 kDa [203, 204]. As previously discussed, wheat gluten consists of a heterogeneous mixture of seed storage proteins namely gliadins and glutenins. The other proteins stored in cereal grains are albumins and globulins [205]. The latter proteins include metabolically active enzymes, their precursors and structural proteins that are present in trace quantities, except for α-/β-amylases and trypsin inhibitors in the albumin fraction and triticins in the globulin fraction [206]. Regardless of the analysis made for the gliadin raw extract, only gluten proteins (gliadins and glutenins) were taking into account. Albumins and globulins are less hydrophobic and are smaller in size than prolamins (they probably represent the proteins on the 10-20 kDa mass range, *Fig. 13*). Since they cumulatively

represent a very small fraction of storage proteins, negligibly influence dough quality parameters, and are of minor nutritional interest then these proteins will not be further discussed.

In order to better differentiate the proteins contained in raw sample and further identify its gliadin components, a bottom-up proteomic approach was followed. *Figure 14* displays the obtained protein bands after SDS-PAGE separation. The resulting protein distribution pattern of the raw sample shows the typical wheat protein components that can be grouped into HMW-GS (65−90 kDa), ω-gliadins (44−55 kDa), LMW-GS (30−40 kDa), and α/β- and γ-gliadins (30−45 kDa). Bands relative to potential albumins/globulins were also detected (10-20 kDa).

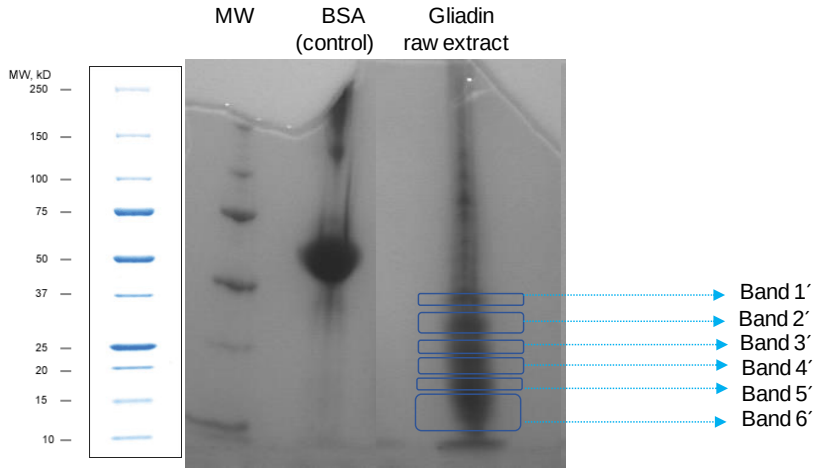


Fig. 14 – SDS-PAGE of the wheat gliadins raw extract using BSA as an internal control. Molecular weight (MW) standards (left) were used as mass references (kDa).

As *Fig. 14* shows, there was a poor separation of the raw sample protein content. Significant differences in bands size and intensity could be observed within the gel. As the most intense ones were embedded between 15 and 37 kDa, then that area was selected to deterrmine those proteins identity. Six consecutive bands were excised and further submitted to tryptic digestion for protein identification.

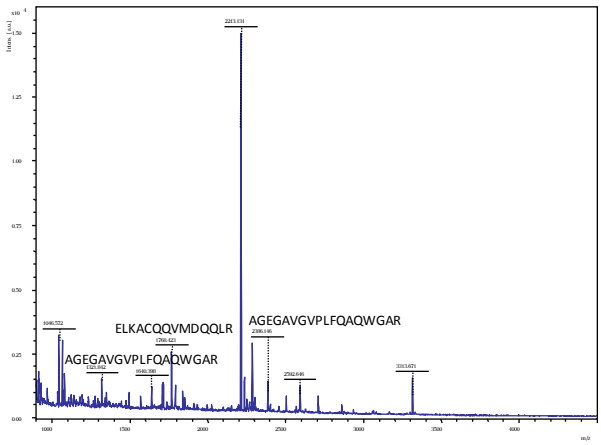
The tryptic digests were first analyzed by LC-ESI-MS/MS using a LTQ-OrbiTrap XL mass analyzer (*see experimental, 2.4.4.1 section*). However, in these conditions, there was no detectable signal. Given the lack of sensitivity of the ESI-MS/MS analysis, a Nano-ESI-MS/MS source was tried (*2.4.4.2 section*). Here, we were able to achieve a good accuracy and sensitivity. After running the obtained data in the Proteome Discoverer 1.4 software, several proteins were proposed to be in the six excised bands. Those results are presented as supplementary information at the end of this dissertation.

Table 6 – Identification of some protein components present in the wheat gliadins raw extract.

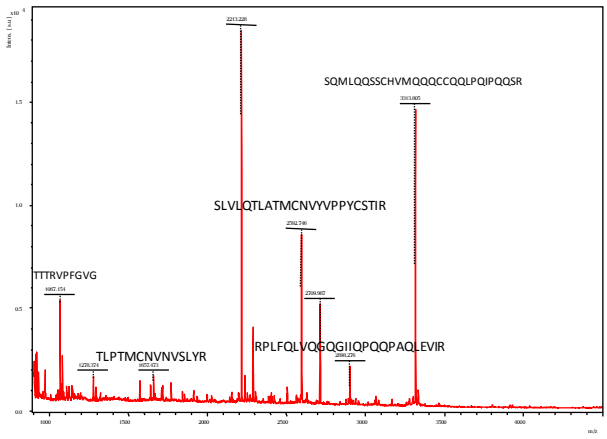
BAND	PROTEIN ID	MW
BAND 1`	LMW-glutenin (Q8W3W8)	35.4 kDa
BAND 2`	Alpha-gliadin (R9XW75)	34.4 kDa
	LMW-glutenin (B1A3G9)	33.8 kDa
	Gamma-gliadin (B5ANT1)	33.2 kDa
	Avenin (AVLB3)	32.4 kDa
BAND 3`	Alpha/beta-gliadin (I0IT55)	33.5 kDa
	Gliadin (R4VEK6)	32.5 KDa
BAND 4`	LMW-glutenin (B8XU58)	31.5 kDa
	Gamma-gliadin (B8XU42)	32.5 kDa
	LMW-glutenin (Q8W3W5 and Q8W3W6)	26.7-29.3 kDa
BAND 5`	Gliadin (R4VEK6)	32.5 kDa
BAND 6`	Gamma-gliadin (Q94G97)	28.9 kDa
	Globulin (Q0Q5D9)	24.5 kDa
	Gamma-gliadin (Q1W676)	14.3 kDa

Except for the bands located above 40 kDa that escaped identification, the remaining protein components (15-37 kDa) were characterized by Nano-ESI-MS/MS. The comprehensive list of most important protein hits given by the SEQUEST search engine is shown in *Table 6*. Only the compounds giving the most intense chromatographic signals were identified. These signals, at times, appeared minimized by contaminant analytes with a high ionization ability. To overcome this cross-contamination during sample preparation and achieve a good accurancy in the mass measurements, some environmental polysiloxanes were used as references in each run.

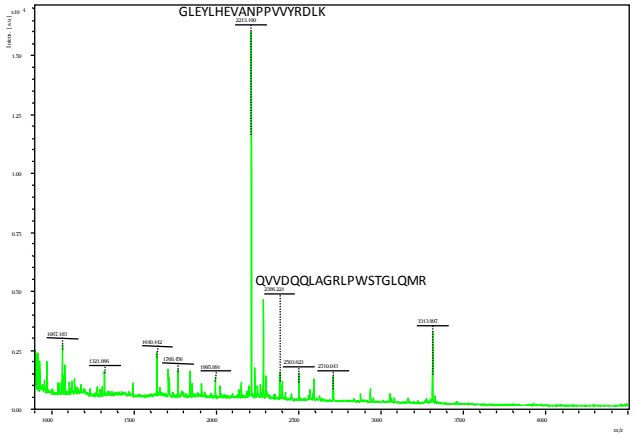
Occasionally, the obtained score values were small. This happened when the signal to noise ratio was low (a match to the "correct" sequence might not exceed this absolute threshold). Sometimes, the number of queries was comparable with the number of entries in the database, meaning that there can be random, low-scoring matches for every entry. Although the average number of random matches per entry may be low, the actual number will follow a distribution, and some entries had large numbers of low scoring matches, resulting in high protein scores. On the other hand, the coverage of most of the identified proteins never reached values higher than 70%. Their poor separation by SDS-PAGE, enhanced by the high sample complexity lead to a low signal to noise ratio for some peptides which difficulted the analysis. Regardless of the unforeseen difficulties, the majority of gluten proteins in sample (from 15 to 37 kDa) could be appropriately identified. In order to confirm the identity of these proteins, MALDI-TOF-MS was tested as an alternative method (*Fig. 15*).



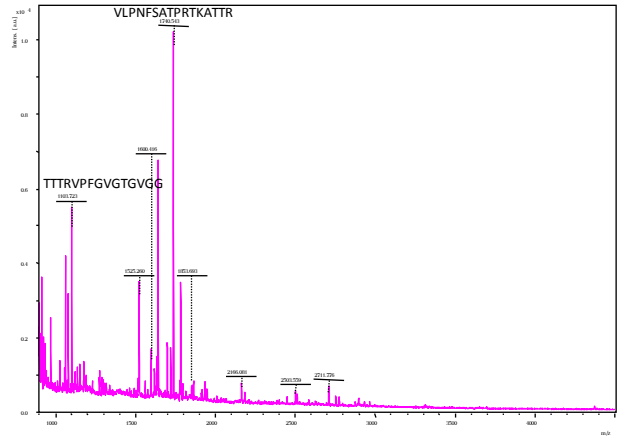
Band 1`



Band 3`



Band 4`



Band 6`

Fig. 15 – MALDI-TOF-MS analysis of tryptic digests.

Figure 15 shows the mass spectra of some tryptic digests (band 1', 3', 4' and 6'). The results presented therein are in agreement with those obtained by ESI-MS/MS (*supplementary information*). The main peptides in each spectrum correspond to the proteins identified in *Table 6*. The poor protein separation by SDS-PAGE is clearly denoted in *Fig. 15* since there are several peptides repeated in contiguous bands.

As it is displayed in *Fig. 16*, thirteen proteins belonging to the wheat gliadins extract were identified following MALDI-TOF-MS, SDS-PAGE and Nano-ESI-MS/MS analysis.

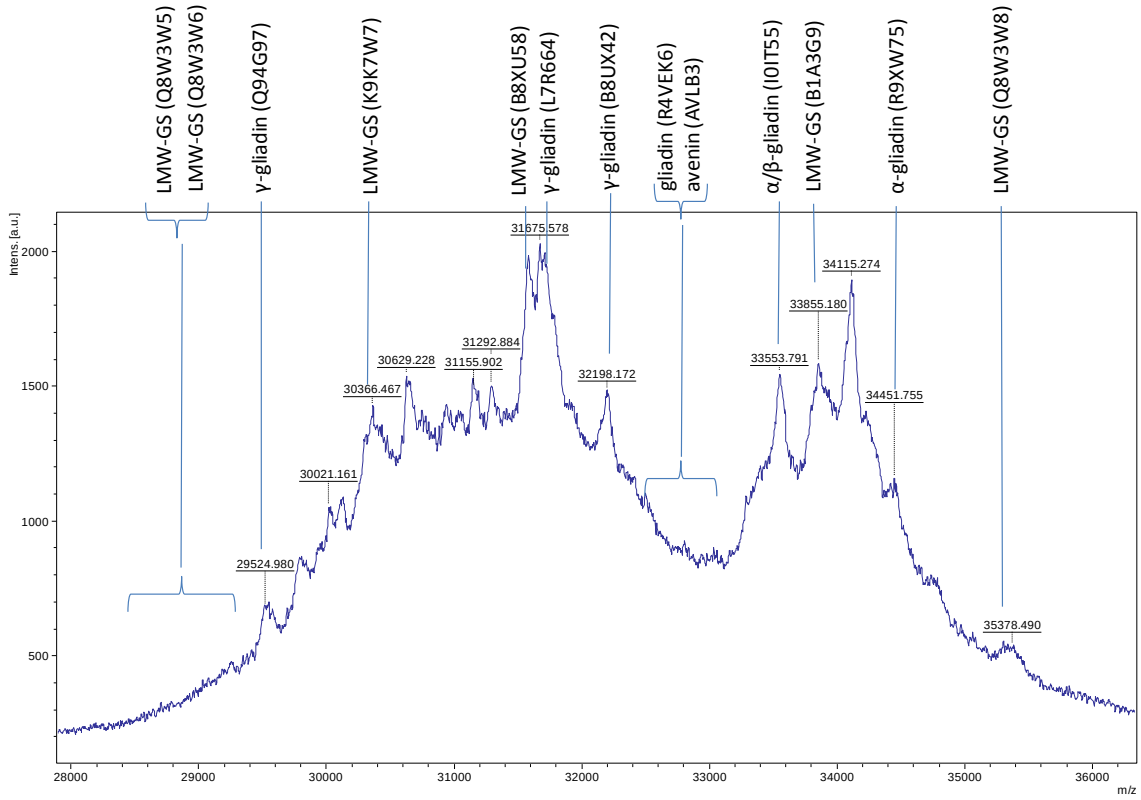


Fig. 16 – Identification of proteins in the wheat gliadin extract by MALDI-TOF-MS, SDS-PAGE and Nano-ESI-MS/MS.

It should to be noticed that the results presented herein only report the identity of the main gluten proteins present in the wheat gliadins raw extract but not all. For that, a further separation involving multistep extraction procedures will be needed. Once separated, that would lead to a better protein analysis and subsequent identification. However, in that context, the main goal of this study would became disrupted. To verify the binding affinity of CD related peptides towards food tannins, it is essential to first study and work with the raw extract as a whole. Future studies will be focused in better understand and differentiate the interaction between different tannins and individual gluten proteins.

3.3 SEPARATION AND CHARACTERIZATION OF WHEAT GLIADIN PEPTIDES

In order to study the ability of different tannins to interact with gliadin-derived peptides, seven peptidic fractions were isolated by semi-preparative HPLC after enzymatic hydrolysis of the wheat gliadins raw extract. *Figure 17* shows the

acquired chromatographic profile of the simulated in vitro digestion. Each fraction was collected in different time intervals, as represented in the following chromatogram.

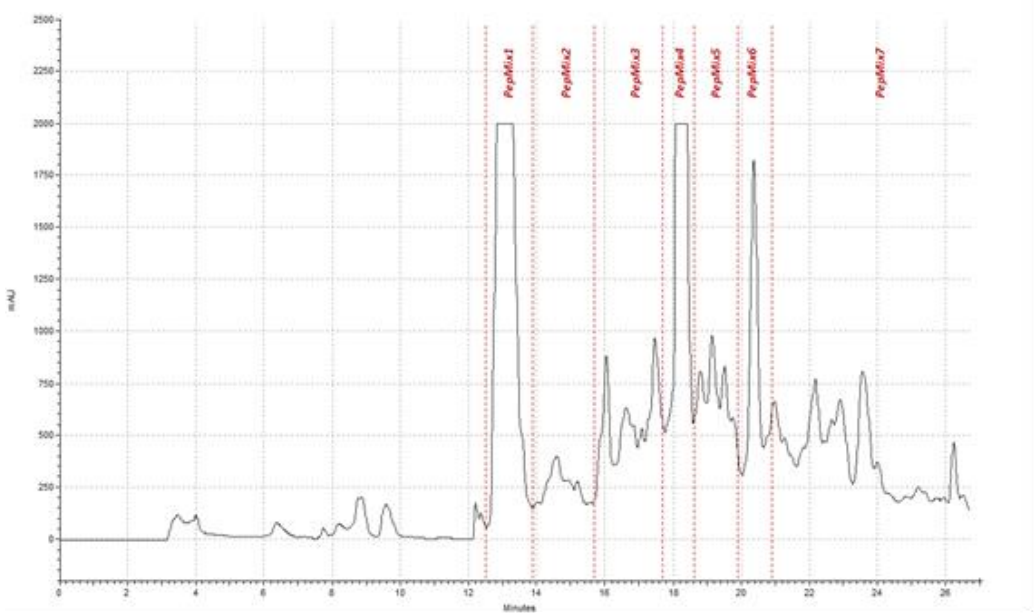


Fig. 17 – Chromatographic profile, acquired by HPLC semi-preparative, of the simulated in vitro digestion of the wheat gliadins raw sample. Each peptide mixture was collected in different time intervals, as represented in the chromatogram.

Thus, *Pep Mix1* corresponded to the first eluted peak (~13 min) while *Pep Mix2* corresponded to the chromatographic zone ranging from 14 to 16 min. *Pep Mix3* relates to the following region (16-18 min) and *Pep Mix4* corresponded to the second major peak (~18 min). *Pep Mix5* matched the chromatographic zone comprised between the second and the third (*Pep Mix6*) major peaks, and finally, *Pep Mix7* corresponded to the subsequent area ranging from 20 to 26 min. Over the past twenty years, mass spectrometry has emerged as a fast and powerful tool in life sciences to determine the identity, quantity, and structural properties of protein and peptide molecules. Therefore, the sequenciation and further identification of the previous wheat gliadin-derived peptide fractions was performed by LC-ESI-MS/MS (*Fig. 18*) in which the amino acid sequences of their main proteolytic products were determined based on those peptides fragmentation pattern. *Table 7* shows the comprehensive list of characterized peptides that were found in the different peptidic fractions resulting from incomplete degradation of gliadins. *Figure 19* shows the number of peptides found in each collected fraction as well as the average number of amino acid residues per peptide. Accordingly, it was observed that the peptides size increased slowly along with their elution order.

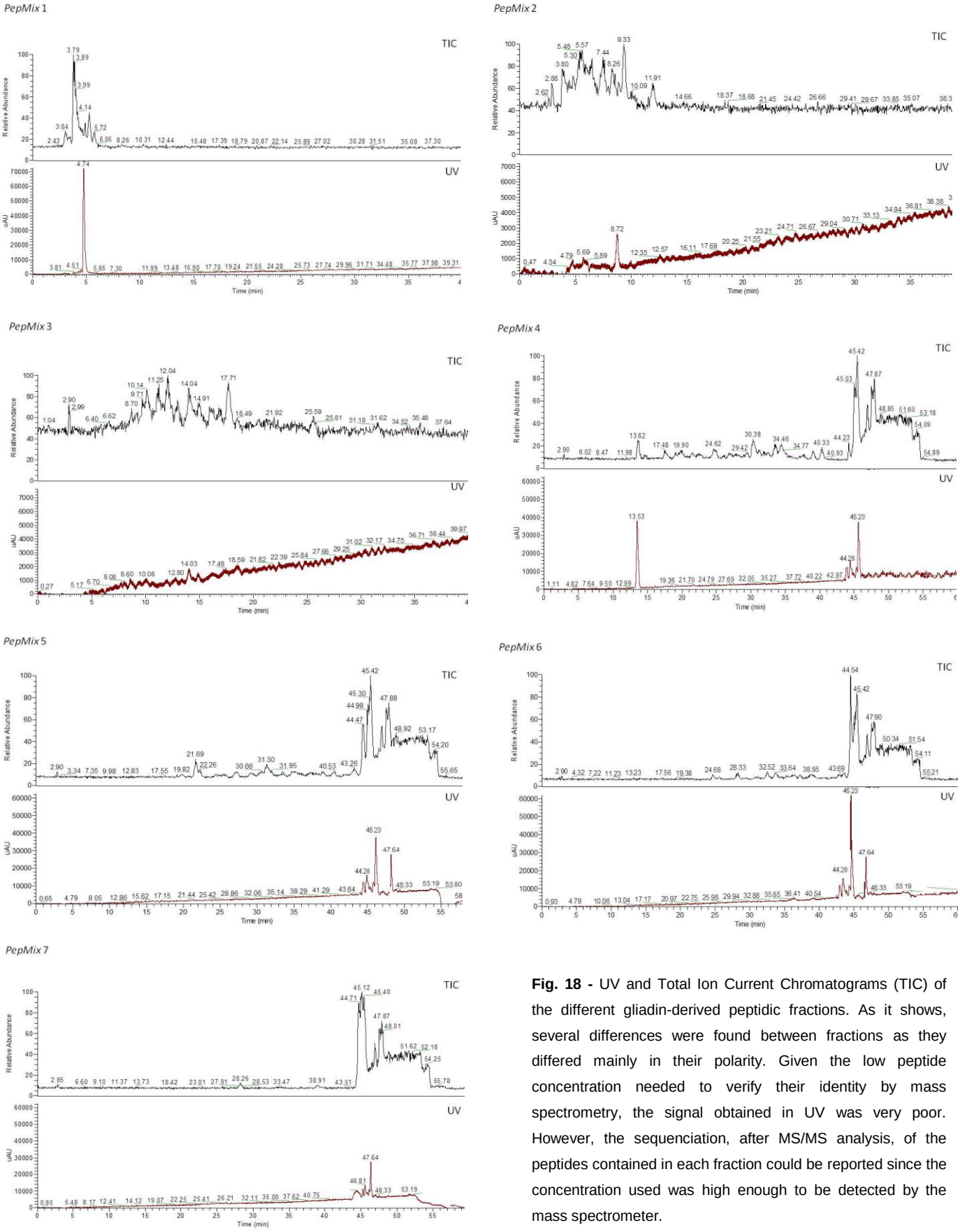


Fig. 18 - UV and Total Ion Current Chromatograms (TIC) of the different gliadin-derived peptidic fractions. As it shows, several differences were found between fractions as they differed mainly in their polarity. Given the low peptide concentration needed to verify their identity by mass spectrometry, the signal obtained in UV was very poor. However, the sequenciation, after MS/MS analysis, of the peptides contained in each fraction could be reported since the concentration used was high enough to be detected by the mass spectrometer.

Table 7 - Identification, by LC-ESI-MS/MS, of peptides generated after enzymatic hydrolysis of the wheat gliadins raw extract. For each collected peptide mixture, different peptide colors and asterisks are indicative of a specific precursor protein. The peptides containing known epitopes that are recognized by intestinal T cells of CD patients are underlined and highlighted in bold.

Pep Mix 1	**PEPTIDES**	GNLGG, AGQGPQ, GQASPQ, TLSASQ, RPGCST, MCSVSV, QQQGLG, RPSQQNPQ, QSRQGQQPGQ
	PROTEINS DERIVED	Alpha/beta-gliadin [M8A3U4_TRIUA]
Pep Mix2	**PEPTIDES**	GQGLPG, TIPSAG, SPVAGQ, ADIGGQ*, GQVPLA, SYDVAG, GSVQPQ, MATTTT, AANMQV**, VLTTLP, GQVLCH, EGRDAL, PQGECQ, QDQQPG, HVSVEH, EHPVPH, PQLQNPS, SQQQQPV*, AQGSVQPQ, GQWHQPGQ, QQEQQGIQ, LSGQGQRPGQ, SHHQQQQPV**, THHQQQQPI*, LQPKNPSQQQPQ
	PROTEINS DERIVED	Alpha/beta-gliadin clone PW8142 [GDA7_WHEAT] *Gamma-gliadin [Q94G97_WHEAT] **Gamma-gliadin [R9XUA5_WHEAT] High molecular weight glutenin subunit 1Dx [S4U495_9POAL] *Low-molecular-weight glutenin subunit (Fragment) [R4JBG2_WHEAT] **Low molecular weight glutenin subunit D3-3 [D3U328_WHEAT]
Pep Mix3	**PEPTIDES**	VGIGGQ**, PPDTGG, DVGTGV**, TYPGGA, AAQLPA*, ALGRAP*, VQGQGI**, VAGIGGQ*, GVGTGVGG*, SIGTGVGG, YPTSPQ*, IPSVATY, GGSFYPGA*, NIQVDPS*, RPQQPY*, QPRQPY*, QDQQSGQ*, NMQVDPS*, SQPQQPI*, SQPQHPI**, YYPTSPQ*, TSIALHNL, QTFPHQPQ*, QPRQPYPQ*, LQELCCQQL*, RGEQHSSCQT, MNIQVDPSSQV**, GSFQPSQQNPQ, LQPQNPSLQQPQ, AEGEASGQLQCERG**, QPRQPFLQPQQP*, SQQPGQGQQGYYSGS**
	PROTEINS DERIVED	*Alpha-gliadin [D2X6D2_TRITD] **Alpha-gliadin protein [X2KVH9_WHEAT] Alpha/beta-gliadin [I0IT51_WHEAT] *Gamma-gliadin [R9XV87_WHEAT] **Gamma-gliadin [Q8L6B3_WHEAT] *HMW-glutenin subunit Glu-1Stx1 [L0G7U5_WHEAT] **High molecular weight glutenin x-type [B8XU61_TRIMO] *Low-molecular-weight glutenin subunit group 3 type II (Fragment) [Q8W3W5_WHEAT] **Low-molecular-weight glutenin subunit (Fragment) [F5A655_WHEAT]
Pep Mix4	**PEPTIDES**	NGAHAI, AIHGVV**, HQAAGL, MATTIA*, GIIQPQ**, VSFQPS, QQQPPS, YPISPQ, VRVPVPQ, QPSHQQP, SHQQQPF, RQPQQPF*, QRPQQPF**, GIIQPQQPA*, LLQQCRPV, YSQPQQPI, HLQPQQPI, QSRCNVMQ, QQPSYSQQ, GQPQQQVPV, LQQQQFPQ, SQQQQPPFS, QQNPQAQGSV**, SHIPGLEKPS, VQGQGIIQPQ, SHIPGLERPS, QQAGQWQRP, SQQQQPVIPQ, DVSPECQPVVV, SQQQPPFSQQ, QQPPQQFPQP**, **TQQPQQPFPQ**, QQGYDNPYHV, EQQPGQGQEGY, ECCSRLGQMPP, QPERGCSGESTA, SFQSSQQNPQA, QSGQGQQGHHSA, SFPQSKQPQQP, SQQQQQPVLPQ, QQPEQGQPGYY, DSPYYVSAEHQ, QQQPGQGQQGHC, QPEQGQEGYYP, VQGQGIIQPQQPA**, PQQYSPYQPQQ, QPIPQPQQPQQP**, SKQPQQPFPQPQ**, **TQQPQQPFPQQP**, SQQQQPVLPQQPS, CSPTPYVQSQMWQ, QQGQQPGQGQQGQQP, SRQQPLPPQQTLSH, **SQQPQQPFPQPQQPQ**, SHIPGLERPSQQQPLPPQQT
	PROTEINS DERIVED	Alpha-gliadin [Q5NDA5_TRITD] *Gamma-gliadin [B6UKP8_WHEAT] **Gamma-gliadin [B6UKQ2_WHEAT] High molecular weight glutenin x-type [B8XU61_TRIMO] LMW-GS [R9XVA5_WHEAT] Gamma-hordein-3 [M7ZUH8_TRIUA]
Pep Mix5	**PEPTIDES**	GQGLPG, ISPAPT, GEGAQQ, NAAHAI, GVNVPL, LGVPIL, KAGSFC, PQCSPV, SSCQSM, SGQGNQG, SQQPPF, QPSILR, MNTFLV, TTTSVPF, QPQCSPV, NPQAQGSV, VQQQLPV, PQLQQPL**, PEQEQQP, SQQPPFPQ, QEQQGQQP, GIIQPQQPA*, QQSQQPFP**, FSQQQQTV, HSKQGQQPG, FQQPHQPF**, YYSTPPQQP, QLLYPQPQP, PQLPCPQPQP, DQCCLQLRPV, **TQQPQQPFPQ****, EQQHGQGLSQP, EQQPGQGQEGY, EQPGQWQQGY, QQGYDNPYHV, SFQSSQQNPQA, QPERGCSGESTA, QQQQPSYSQQ, QQQIPVVQPSV, SQQAFPQPQQT, PGQGQQGHCPASQ, AATPNSSTEVMTS, AQMETNCISGLE, QQQPGQGQQGHC, QPEQGQEGYYP, GGSFYPGETTPPQ, VQGQGIIQPQQPA**, VQGQAIIQPQQPA**, **SQQPQQPFPQQP**, **TQQPQQPFPQQP****, **LQQPQQPFPQPQ****, LQPGQGQQGYYPT, QSISQYQQQQPQ, SQQQQPVLPQQPP, YQQPQQTFPQPQ, QPGYYPTSPWQPE, ILQQQQQQQQQQV, QCAAINSVVHSIMMQQE, LQPQNPSQQQPQEQVPL, **TQQPQQPFPQQPQQPFPQ****, SHIPGLERPSQQQPLPPQQT, SQHQQQQQQQQQQQQQQQI
	PROTEINS DERIVED	Alpha-gliadin [K7WV57_WHEAT] *Gamma-gliadin [R9XSZ2_WHEAT] **Gamma-gliadin [Q9M6P7_WHEAT] Low-molecular-weight glutenin subunit [Q1ZZT4_WHEAT] Putative puroindoline-like protein [B1GXL7_TRITD]
PepMix 6	**PEPTIDES**	GIGVGV, GGARGL, LSGAIP, DPLGAL, GEGAQQ, APSGIF, GGGGGSQQ, GQVQEP, QQGTFL, SIILPR, SQVCFQ, ALETLPA, HVNTEQ, VQQGTFL, QPSQQNP, KKITLVL, CQQQPQP, GQVQWPQ, QIPEQSR, QQSSYQVL, QLEMMTSI, QQPHQAFPQ, LQQPGREQQ, QQPLQQYPL, QEPGQGQQWY, QEPGQWQQGY, SHIPSLEKPL, EQQPGQGQEGY, SFQSSQQNPQA, VPQPQQPQQPF, PFCQHQQPFY, MQQQCCQQLW, GQPGCDPTSPQQPG, MWQRSSCNVMQ, SSSPQQLGQGQPRY, HPSILQQLDPCKVF, LQPQNPSQQQPQEQVPL, SQQQQSPFSQQQQQQPPF, **PQQPFPQQPQQPYPQQPL**, **TQQPQQPFPQQPQQPFPQ**, **MQLQPFPQPQLPYPQPQL**
	PROTEINS DERIVED	Alpha-gliadin [R9XV48_WHEAT] High-molecular-weight glutenin [Q41516_WHEAT] Gamma-gliadin [Q9M6P7_WHEAT]
Pep Mix7	**PEPTIDES**	AVGVPV, VSSKAG, GGGGGSQQ, SQHEQV, QGSVQPK, PSRQQPQ, VPGLEKPW, KANLRLATMK, LQLQPFPQPQ, QQGYDSPCHVS, RNNNSPGHNNP, QQGYDNPYHV, EQQPGQGQEGY, YSSTTSMPFSI, QEPGQWQQGY, QQGYDSPCHVS, QQGYDNPYHV, QSRCNVMQQQ, QQQQPSYSQQ, QPERGCSGESTA, QQPEQGQPGYY, QPGQEQQGYYP, QPEQGQQGYYP, QQQPGQGQQGHC, QPEQGQEGYYP, MWQRSSCNVMQ, EHYTPCMTFLQ, GLRVGGNQIAGHIP, LQLQPFPQPQLP, SQQQQPVLPQQPPFS, **LQPQQPFPQQPQQPFPQ****, LQLQPFFPQPQPFPPQLPYPQPQ, **TQQPQQPFPQQPQQPFPQQPQQPFPQ***, **QQPFPQQPQQPQQPFSQPQQQLPLQPQQPFP**
	PROTEINS DERIVED	Alpha-gliadin [Q27HK6_TRIMO] Alpha/beta-gliadin MM1 [M7ZZV2_TRIUA] *Gamma-gliadin [R9XUS6_WHEAT] **Gamma-gliadin (Fragment) [Q9M6P7_WHEAT]

Although the seven peptidic fractions were mainly composed from six up to eleven amino acid residues, longer peptides with more than seventeen amino acids appeared from *Pep Mix4* to *Pep Mix7*. Indeed, the higher peptide (31-mer) was sequenced in *Pep Mix7*.

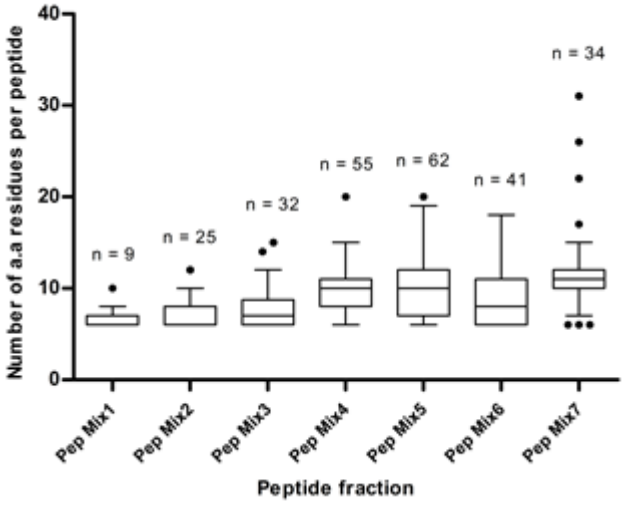


Fig. 19 – Number (n) and size distribution (average number of amino acid residues per peptide) of gliadin-derived peptides, obtained after *in vitro* digestion of the wheat gliadins crude extract and further fractionation. It displays the median and Tukey test of the obtained values.

From all the structurally characterized peptides, only a small percentage were truly identified as wheat gliadins or even glutenins digestive products (*Fig. 20*). As aforementioned, this feature might be due to the occurrence of other protein components in raw sample that were not considered in data treatment given their potential irrelevance on a CD point of view (non-gluten proteins). Indeed, as previously described (*see experimental*), only gluten proteins were selected to run the identification software process. Interestingly, the majority of the unidentified ones appeared to be concentrated in the last four peptidic fractions where they represented more than half of their respective content.

Additionally, several differences in the distribution of gluten protein subtypes were observed between fractions (*Fig. 20*). In that way, while α/β-gliadin derived peptides were found in ascendant order of appearance in *Pep Mix1*, *2*, *7* and *3*, the α-gliadin derivatives were mainly found in *Pep Mix6* followed by *Pep Mix7*, *3*, *5* and *4*. The γ-gliadin derivatives were mainly present in *Pep Mix4* and *5* followed by *Pep Mix3*, *7*, *2* and *6*. High-molecular-weight (HMW) glutenin derivatives decreased from *Pep Mix2* to *4* although some peptides were also present in *Pep Mix6*. Low-molecular-weight (LMW) glutenin-derived peptides increase their quantity from *Pep Mix2* to *Pep Mix5*.

Overall, it should to be noted that the distribution of peptides belonging to a specific precursor protein is not randomized: while the α/β gliadin-derived peptides were primarily embedded in the first three fractions, the glutenin-derived peptides were mainly found in intermediate peptidic fractions (from *Pep Mix2* to *Pep Mix5*). γ-gliadin derived peptides started their elution in *Pep Mix2* reaching a maximum in *Pep Mix4* and then decreased until the last fraction. The α-gliadin derived peptides appeared mainly in the last two fractions (*Pep Mix6* and *7*) despite their elution started in *Pep Mix3*.

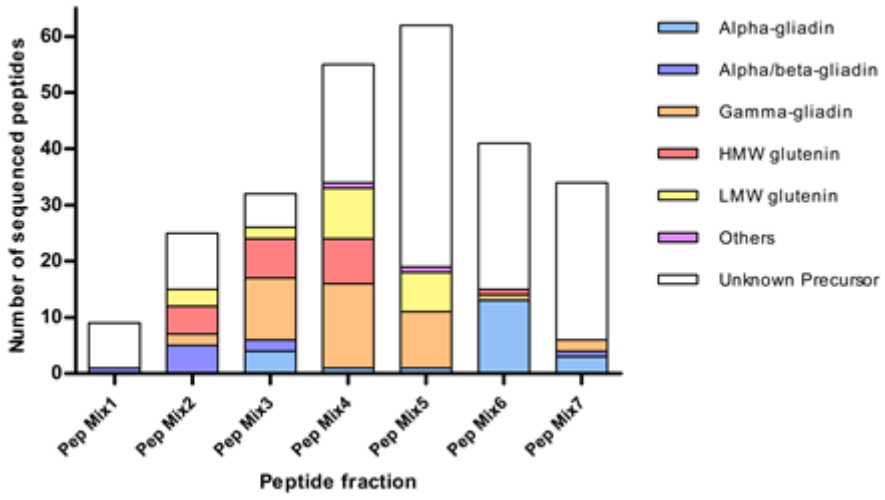


Fig. 20 – Number and precursor protein distribution of all sequenced peptides in each collected fraction.

Besides size and composition, some CD relevant differences were also observed among each collected peptide fraction. In that way, five distinct CD T cell epitopes [28, 62, 68, 69, 207] were found from *Pep Mix4* to *Pep Mix7* (*Table 8*): QQPQQPFPQ (*a-type*), QQPFPQQPQ (*b-type*), QQPQQPYPQ (*c-type*), PFPQPQLPY (*d-type*) and PQPQLPYPQ (*e-type*). This nomenclature (*a-* to *e-type*) was adapted in order to simplify the presented results.

Pep Mix4 presented only the *a-type* epitope in the peptides TQQPQQPFPQ, TQQPQQPFPQQP and SQQPQQPFPQPQQPQ; *Pep Mix5* presented the *a-type* and *b-type* epitopes in the peptides TQQPQQPFPQ (*1x a-type*), TQQPQQPFPQQP (*1x a-type*), LQQPQQPFPQPQ (*1x a-type*), SQQPQQPFPQQP (*1x a-type*) and TQQPQQPFPQQPQQPFPQ (*1x a-type* and *b*-type); *Pep Mix6* presented the *a-type*, *b-type*, *c-type*, *d-type* and *e-type* epitopes in the peptides PQQPFPQQPQQPYPQQPL (*1x b-type* and *c-type*), TQQPQQPFPQQPQQPFPQ (*1x a-type* and *b-type*) and MQLQPFPQPQLPYPQPQL *(1x d-type and e-type)* while *Pep Mix7* presented the *a-type* and *b-type* epitopes in the peptides TQQPQQPFPQQPQQPFPQQPQQPFPQ (*2x a-type* and *1x b-type*), QQPFPQQPQQPQQPFSQPQQQLPLQPQQPFP (*1x b-type*)

and LQPQQPFPQQPQQPFPQ *(1x a-type and b-type)*. Therefore, focusing on the potential relevance of the peptides contained in each fraction, it should be noticed that not only their size but also their content in previously described CD epitopes could distinguish one fraction from the others. According to *Sollid 2002* the minimum length required for T cell recognition is fixed in nine amino acid residues [16]. Hence, *Pep Mix1*, *2* and *3* could be the less relevant since their peptides were mainly composed by six amino acid residues and also they did not contain any CD epitope. Oppositely, *Pep Mix4*, *5*, *6* and *7* should be, at first, the most important ones because their peptide content accomplished the requirements in terms of size and CD epitopes.

Table 8 – Sequenced peptides containing CD epitopes (identified in different colors) [28, 62, 68, 69, 207] as well as their precursor protein code and name.

PEPTIDE FRACTION	PEPTIDES CONTAINING CD EPITOPES	CODE	PROTEIN DERIVED
Pep Mix4	TQQPQQPFPQ (1x QQPQQPFPQ)	B6UKQ2	γ-gliadin
	TQQPQQPFPQQP (1x QQPQQPFPQ)	B6UKQ2	γ-gliadin
	SQQPQQPFPQPQQPQ (1x QQPQQPFPQ)	B6UKQ2	γ-gliadin
Pep Mix5	TQQPQQPFPQ (1x QQPQQPFPQ)	Q9M6P7	γ-gliadin
	TQQPQQPFPQQP (1x QQPQQPFPQ)	Q9M6P7	γ-gliadin
	LQQPQQPFPQPQ (1x QQPQQPFPQ)	Q9M6P7	γ-gliadin
	SQQPQQPFPQQP (1x QQPQQPFPQ)	-	-
	TQQPQQPFPQQPQQPFPQ (1x QQPFPQQPQ) (1x QQPQQPFPQ)	Q9M6P7	γ-gliadin
Pep Mix6	PQQPFPQQPQQPYPQQPL (1x QQPFPQQPQ) (1x QQPQQPYPQ)	-	-
	TQQPQQPFPQQPQQPFPQ (1x QQPFPQQPQ) (1x QQPQQPFPQ)	-	-
	MQLQPFPQPQLPYPQPQL (1x PFPQPQLPY) (1x PQPQLPYPQ)	R9XV48	α-gliadin
Pep Mix7	TQQPQQPFPQQPQQPFPQQPQQPFPQ (1x QQPFPQQPQ) (2x QQPQQPFPQ)	R9XUS6	γ-gliadin
	QQPFPQQPQQPQQPFSQPQQQLPLQPQQPFP (1x QQPFPQQPQ)	-	-
	LQPQQPFPQQPQQPFPQ (1x QQPFPQQPQ) (1x QQPQQPFPQ)	Q9MP6P7	γ-gliadin

3.4 FLUORESCENCE QUENCHING STUDIES

For fluorescence quenching measurements, only *Pep Mix4* and *Pep Mix6* were used since they were the ones who presented, in these experimental conditions, considerable fluorescence. *Figure 21* shows the fluorescence emission spectra obtained for *Pep Mix4* with the addition of increasing concentrations of different tannin procyanidins (procyanidin B3, trimer T1, tetramer TT1 and FV of oligomeric procyanidins). Independently of the tested tannin it was observed that in all cases their addition caused a gradual decrease in the fluorescence intensity by quenching, without any significant shift in the emission maximum wavelength. The same behavior was observed for *Pep Mix6* (*Fig. 22*).

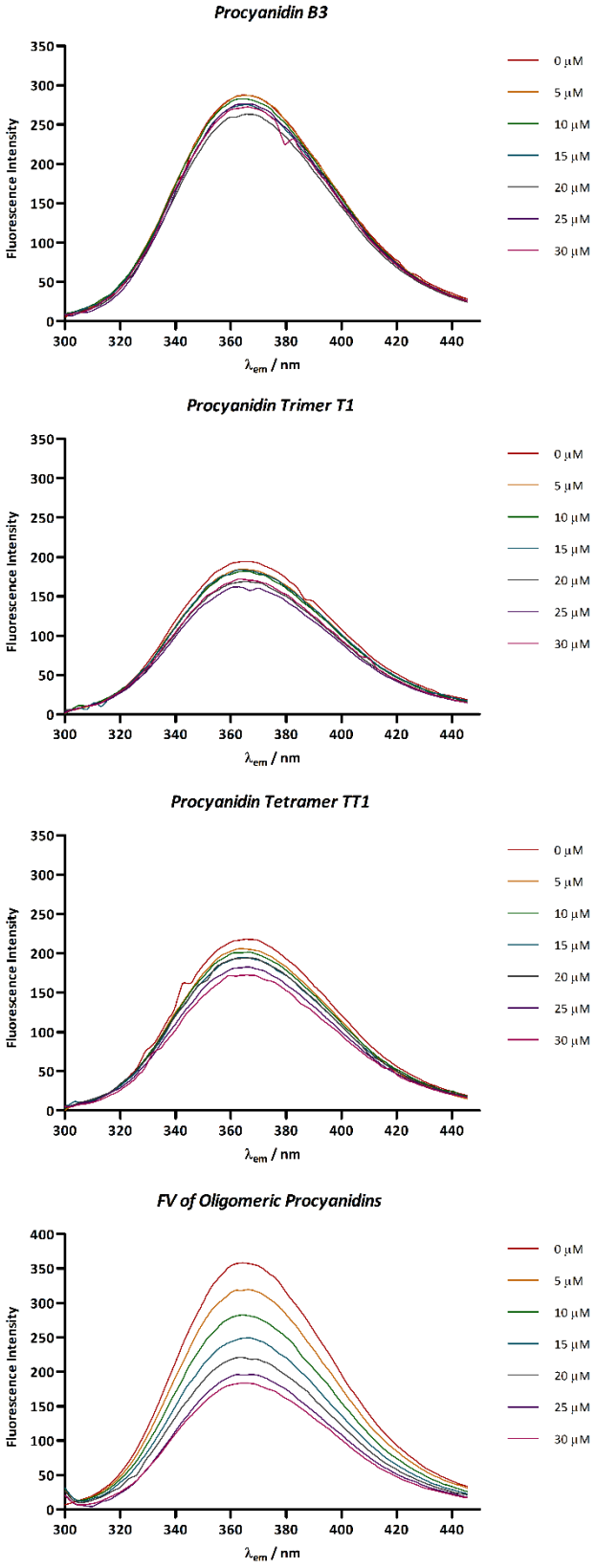


Fig. 21 – Fluorescence emission spectra (at λ_{em} = 290 nm) of *Pep Mix4* (0.02 mg/mL) in the presence of increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FV of oligomeric procyanidins. Each curve represents a triplicate assay after correction for polyphenol fluorescence.

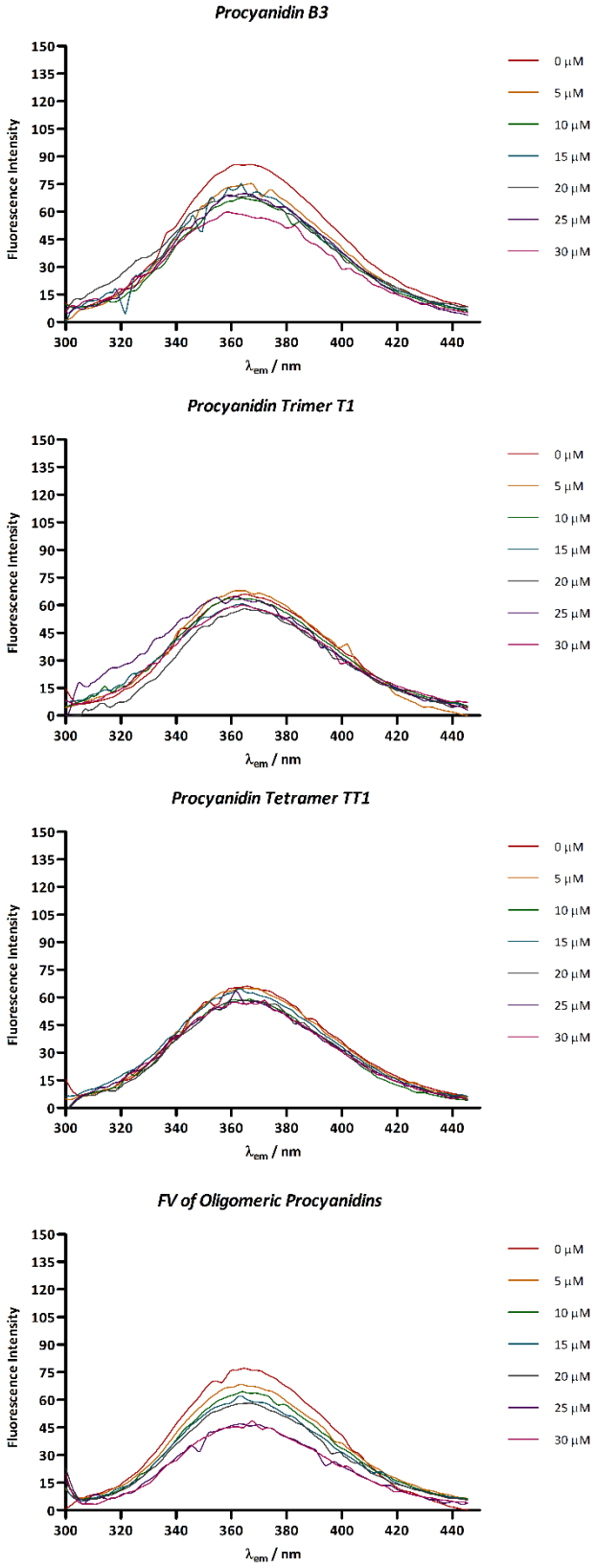


Fig. 22 – Fluorescence emission spectra (at λ_{em} = 290 nm) of *Pep Mix6* (0.02 mg/mL) in the presence of increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FV of oligomeric procyanidins. Each curve represents a triplicate assay after correction for polyphenol fluorescence.

The calculation of K_{sv} from Stern-Volmer plots (*Fig. 23A* for *Pep Mix4* and *Fig. 23B* for *Pep Mix6*) demonstrated, mainly for *Pep Mix4*, that quenching depends on the polyphenolic structure, with fluorescence extinction being determined, in magnitude, by the procyanidins polymerization degree (*Table 9*). The Stern-Volmer constant (K_{sv}) is a measure of the ability of the tested polyphenols to interact with peptides in solution, reducing the fluorescence of the amino acid residue that is fluorescing (in this case tryptophan) and is determined as the slope of the $F_0/F = f([Quencher])$ plot, where F_0 and F are the fluorescence intensities before and after the addition of the quencher. Oligomeric procyanidins of fraction V revealed the highest quenching constant in both peptide mixtures. On the other hand, there were no statistically significant differences between the quenching constants for procyanidin B3, procyanidin trimer T1 and procyanidin tetramer TT1 in *Pep Mix6*. In general, all the studied procyanidins appeared to be slightly more reactive towards the fourth peptide mixture than towards the sixth.

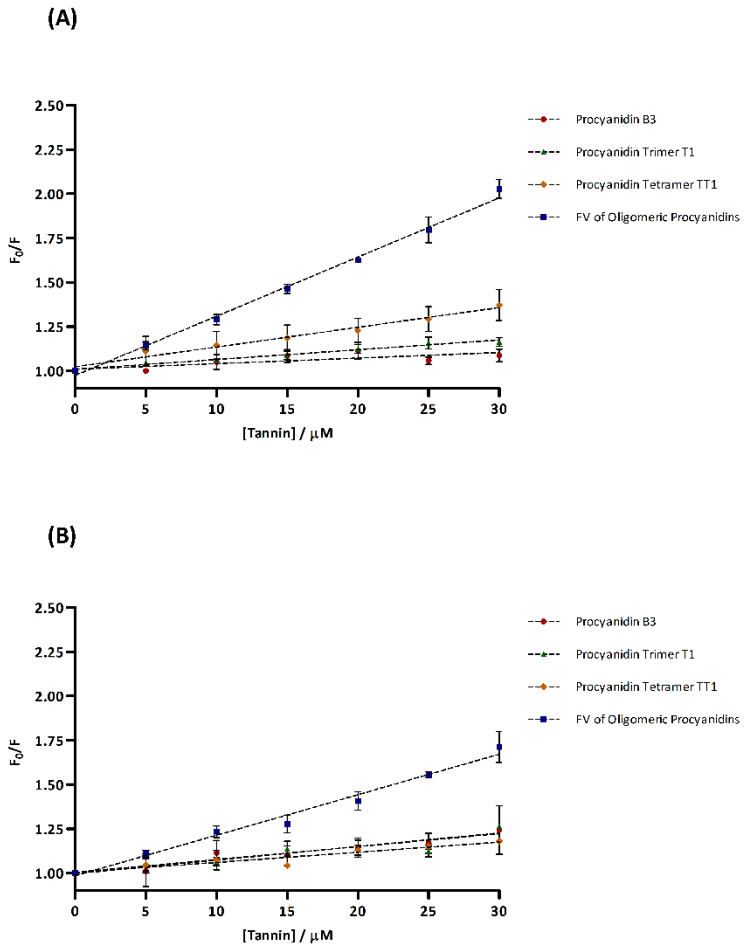


Fig. 23 – Stern-Volmer plots describing tryptophan quenching of *Pep Mix4* (A) and *Pep Mix6* (B) by increasing concentrations of procyanidin B3, procyanidin trimer T1, procyanidin tetramer TT1 and FV of oligomeric procyanidins. The fluorescence emission intensity was recorded at λ_{ex} 290 nm, and the λ_{em} maximum occurred at 365 nm.

Table 9 – Stern-Volmer Quenching Constants (K_{sv}) for the interaction between both *Pep Mix4* and *Pep Mix6* and procyanidins with increasing degree of polymerization (B3, trimer T1, tetramer TT1 and FV of oligomeric procyanidins). Values with different letters (a-e) are significantly different (P < 0.05).

	K_{sv} (M^{-1})			
	PROCYANIDIN B3	**PROCYANIDIN TRIMER T1**	**PROCYANIDIN TETRAMER TT1**	**FV OF OLIGOMERIC PROCYANIDINS**
Pep Mix4	3148 ± 841.3[a]	5543 ± 428.2[a]	11150 ± 1406[b]	33410 ± 906.5[c]
Pep Mix6	7660 ± 1502[b]	7260 ± 946.8[d]	5738 ± 695.7[d]	22890 ± 1089[e]

To interpret the data from fluorescence quenching studies, it is important to understand what kind of interactions take place between the quencher and the fluorophore. As represented in *Fig. 23A* and *Fig. 23B*, for both peptide mixtures and procyanidins tested, the respective Stern-Volmer plots were all linear, which means that only one type of quenching occurred (dynamic or static). In the case of a dynamic mechanism for fluorescence quenching, it is the diffusion-limited collision between the quencher and the fluorophore molecules that allows the energy transfer without radiation. Quenching can also be caused by the formation of a complex between those two compounds that does not fluoresce after returning from the excited state – static quenching [208]. To verify if that quenching is due to a specific interaction, or complex formation, it is essential to calculate the bimolecular quenching constant (k_q), dividing the obtained K_{sv} values by the lifetime of each peptide mixture in the absence of the quencher (τ_0). The dynamic mechanism (diffusion-limited quenching) typically results in values of k_q near 10^{10} $M^{-1}.s^{-1}$ [209]. Since all the obtained values for k_q were more than 90-fold higher, this suggests that the interaction of procyanidins with both peptide mixtures involved the formation of a stable complex (*Table 10*) [210, 211]. In vivo, this kind of interaction could eventually predict the potential of procyanidins to interfere with the availability of celiac reactive peptides, blocking their immunological and toxic effects on the intestinal mucosa. On the other hand, this hypothetical ability of procyanidins to snatch such peptides appears to be highly dependent on its structure, or more specifically its degree of polymerization. In fact, assuming that the Stern-Volmer quenching constant corresponds to a binding constant, this trend is well confirmed in *Fig. 23A*, where its value increased from procyanidin B3 to FV. The essence of such behavior is based on the fact that the number of catechin units and galloyl groups increases with the molecular weight of the procyanidins. This results in a higher number of aromatic rings and hydroxyl groups that may be involved in hydrophobic and hydrogen bonding with several protein binding sites [194]. Therefore, a stronger binding affinity was anticipated for the high molecular weight procyanidin

oligomers, as is the case of FV. For *Pep Mix6*, however, this behavior does not appear to be as linear as previously described, since the constants of procyanidin B3, procyanidin trimer T1 and procyanidin tetramer TT1 showed a similar magnitude (statistically, they are not significantly different) (*Table 9*). The differences between the bindings of the same polyphenol to *Pep Mix4* and *Pep Mix6* may reflect structural differences between those two peptide mixtures, including the amino acid composition of the peptides that were involved in complex formation [147, 212].

Table 10 – Bimolecular Quenching Constants (kq) for the interaction between both *Pep Mix4* and *Pep Mix6* and procyanidins with increasing degree of polymerization (B3, trimer T1, tetramer TT1 and FV of oligomeric procyanidins). Values with different letters (a-e) are significantly different (P < 0.05).

	τ_0 (s)	*kq* x 10^{-12} (M^{-1} s^{-1})			
		PROCYANIDIN B3	PROCYANIDIN TRIMER T1	PROCYANIDIN TETRAMER TT1	FV OF OLIGOMERIC PROCYANIDINS
Pep Mix4	3.509 x 10^{-9}	0.897 ± 0.240[a]	1.580 ± 0.122[a]	3.177 ± 0.401[b]	9.521 ± 0.258[c]
Pep Mix6	3.302 x10^{-9}	2.320 ± 0.455[d]	2.199 ± 0.287[d]	1.738 ± 0.211[d]	6.932 ± 0.330[e]

3.5 LIGHT SCATTERING STUDIES

On a first attempt, it has been tested the possibility of using in the DLS experiments the same tannin concentration range with respect to the fluorescence quenching assays. However, in these conditions, no significant aggregation was detected when different concentrations of procyanidin B3 and oligomeric procyanidins were added to 0.02 mg/mL of *Pep Mix6* (*Fig. 24*). In addition to the measurements being clearly performed below the instrument detection limit, the quality of the solvent also displayed an important influence in the data consistency and variability. In fact two different waters were tested in the DLS assays: a regular distillated water and a Milli-Q ultrapure water. As expected, the best results were obtained using Milli-Q ultrapure water as solvent especially if previously filtered through 0.2 µm disposable PTFE filters. Procyanidin B3 assay was developed in a regular distillated water whereas the other tannins were dissolved in Milli-Q ultrapure water prior to analysis. In *Fig. 24* it is clearly visible the differences between the lower accuracy from procyannidin B3 assay to the other tested tannins. Indeed, the obtained relative standard error (% RSD), a measure of the statistical estimate's reliability, was higher than 20%. There was, therefore, a low precision and repeatability of this assay. Considering that there were not found size differences regardless of the tannin concentration (there was no complexation), the test

was not repeated in Milli-Q ultrapure water regardless on the high standard deviation between replicates found for regular distillated water.

In order to surpass the substantial lower sensitivity of dynamic light scattering compared to florescence quenching, it was decided to increase not only the initial concentration of *Pep Mix6* (0.2 mg/mL) but also the tannin concentrations range up to 500 µM (*Fig. 24*). In these circumstances, a differential aggregation was detected as it was dependent on the polymerization degree of the tested tannin. Indeed, the addition of increasing concentrations of the procyanidin tetramer TT1 resulted in the formation of measurable and successively bigger peptide-tannin aggregates while no significant differences were visible in the presence or absence of procyanidin dimer B3 (*Fig. 24*). Therefore, for the subsequent assays involving all the seven collected peptide mixtures, the concentration of those was fixed in 0.2 mg/mL while the tannin concentration range was increased to 1 mM in order to better evidence the aggregation process.

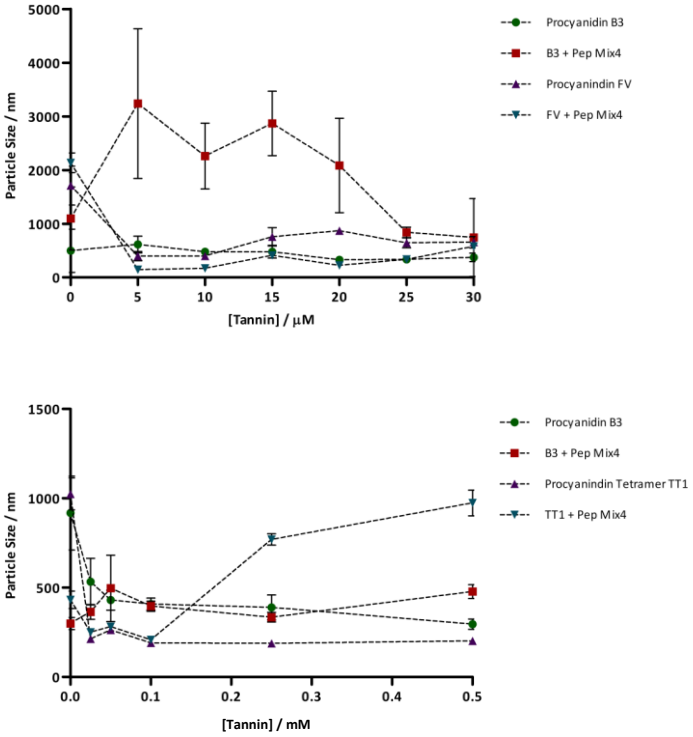


Fig. 24 – Differences in aggregation size of tannin-*Pep Mix4* complexes depending on the peptide concentration (A, 0.02 mg/mL and B 0.2 mg/mL) and as the tannin tested (procyanidin B3 and procyanidin tetramer TT1).

After the experimental parameters were set, light scattering measurements were developed in order to characterize the size of the different aggregates formed between the digested peptides and an oligomeric mixture of tannin procyanidins (FV) (*Fig. 25*). According to DLS, *Pep Mix6* produced the larger aggregates, being this behavior observed across the whole range of tannin

concentration. Still, by visual examination, it was found that the addition of increasing concentrations of procyanidins to *Pep Mix7* resulted in immediate cloudiness and subsequent precipitation of insoluble aggregates in the DLS cell. As an unusual extensive aggregation took place in the latter mixture, it became virtually impossible to correctly measure the aggregates size by such a light scattering study, with the obtained values exhibiting a tremendous variability (*data not shown*). *Pep Mix1* seems to be the less reactive towards oligomeric procyanidins since the dimension of the resulting aggregates, if any, remained nearly unchanged. For the intermediate peptide mixtures (*Pep Mix2* to *Pep Mix5*), no significant differences were detected in the size of the formed aggregates, all of them having a very similar behavior in the whole range of procyanidin FV concentration.

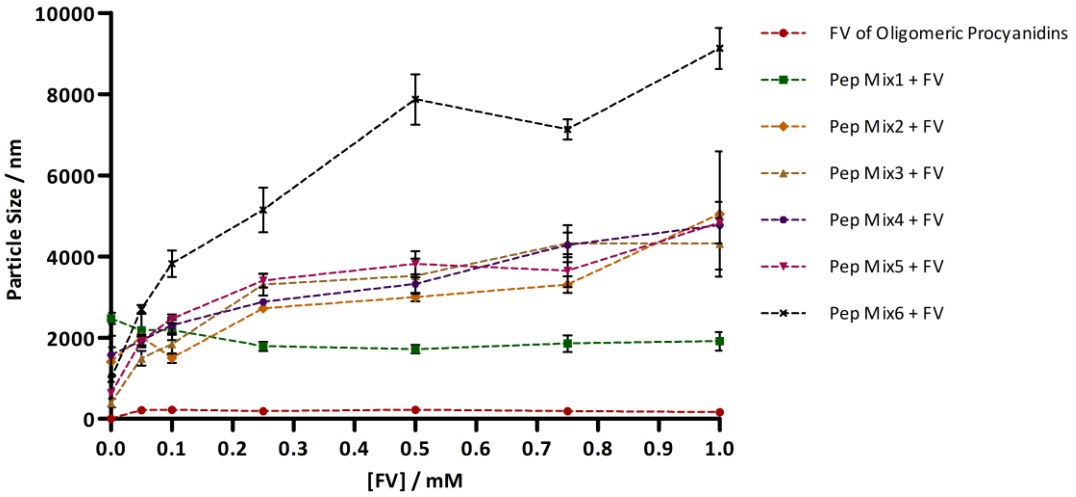


Fig. 25 – Changes in the aggregate size of all the seven collected peptide mixtures (0.2 mg/mL) by increasing concentrations of FV containing oligomeric procyanidins with a high degree of polymerization.

As can be seen, by DLS, the results are somewhat opposed to the ones obtained by fluorescence quenching in that the FV of oligomeric procyanidins appeared to be slightly more reactive towards *Pep Mix6*. The reason for this may be explained based on differences that are inherent to these two techniques. On one hand, fluorescence quenching is an extremely sensitive and selective approach that gives information about the molecular environment in the vicinity of a chromophore molecule. On the other hand, DLS is less selective than fluorescence quenching assays giving information about the size of structures in solution at a supramolecular level. In addition, it is important to highlight that the specificity of tannin-protein interaction, among other things, is strongly

dependent on the protein and tannin concentration range [165, 166]. So, while in dilute solutions the tannins may specifically bind to individual peptides in a way that is essentially determined by both structural features and without any protein aggregation, when the tannin/peptide molar ratio exceeds a threshold, the aggregation of peptides may occur with tannins bridging them together [213]. Since this event is highly favoured by the complexity of the peptides available to interaction, and because the number of peptides with increasing size increased from *Pep Mix1* to *Pep Mix7*, it was assumed, for the peptide/procyanidin concentration range used in DLS, that the size factor may become a much more decisive driving force when determining the dimension of the resulting aggregates. Consequently, peptide mixtures collected later by semi-preparative HPLC produced the largest aggregates eventually leading to their precipitation when the transfer of nonaggregated peptides to the aggregates became too pronounced.

3.6 CHARACTERIZATION OF PROCYANIDIN B3-GLIADIN DERIVED PEPTIDE COMPLEXES

A tannin-peptide binding assay by ESI-MS was developed in order to identify the non-covalent complexes formed between procyanidin B3 and the previously characterized peptide fractions. Initially, mass spectrometry experiments were performed using procyanidin B3 and the synthetic immunodominant CRP32 peptide. As shown in the mass spectrum displayed in *Fig. 26* (and complemented by its legend), after B3-CRP32 reaction, twenty-four new analytes were detected. They corresponded to only four different B3-CRP32 non-covalent complexes along with their respective adducts (mainly Na^+ and K^+ adducts and also NH_4^+ and H_3O^+ losses). The identified complexes had a molecular weight comprised between 2000 and 6000 Da which corresponded to one peptide unit associated with one up to four B3 moieties. Thus, the m/z peaks codified in *Fig. 26* from 10 to 15 and from 20 up to 25 belonged to a B3-CRP32 complex, from 26 to 31 to a 2B3-CRP32 complex and from 32 up to 35 to a 4B3-CRP32 complex. The last two m/z peaks are those representatives from a 3B3-CRP32 complex. It should to be noted that, in most of the cases, the formed complexes were multiply charged so the m/z values showed in the spectra do not correspond to the molecular weight of the analytes.

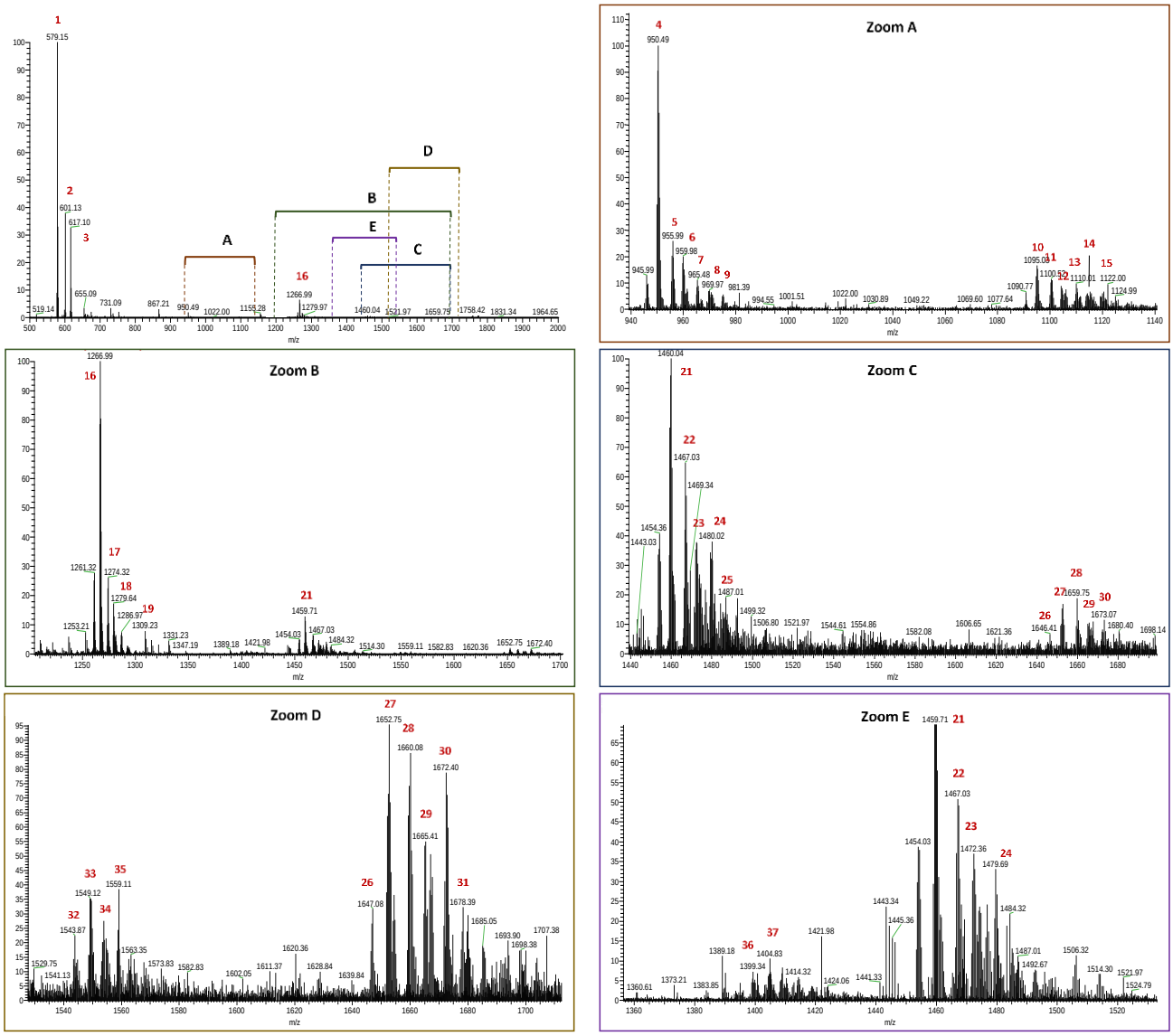


Fig. 26 – ESI-MS spectra with all detected B3-CRP32 complexes. *1*. $[B3+H]^+$, *2*. $[B3 + Na^+]^+$, *3*. $[B3+K^+]^+$, *4*. $[CRP32 + 4H^+]^{+4}$, *5*. $[CRP32 + Na^+ + 3H^+]^{+4}$, *6*. $[CRP32 + K^+ + 3H^+]^{+4}$, *7*. $[CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *8*. $[CRP32 + 2Na^+ + K^+ + H^+]^{+4}$, *9*. $[CRP32 + Na^+ + 2K^+ + H^+]^{+4}$, *10*. $[B3\text{-}CRP32 + 4H^+]^{+4}$, *11*. $[B3\text{-}CRP32 + Na^+ + 3H^+]^{+4}$, *12*. $[B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *13*. $[B3\text{-}CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *14*. $[B3\text{-}CRP32 + 2Na^+ + K^+ + H^+]^{+4}$, *15*. $[B3\text{-}CRP32 + Na^+ + 2K^+ + H^+]^{+4}$, *16*. $[CRP32 + 3H^+]^{+3}$, *17*. $[CRP32 + Na^+ + 2H^+]^{+3}$, *18*. $[CRP32 + K^+ + 2H^+]^{+3}$, *19*. $[CRP32 + Na^+ + K^+ + H^+]^{+3}$, *20*. $[B3\text{-}CRP32 + 3H - NH_4^+ + 4H^+]^{+3}$, *21*. $[B3\text{-}CRP32 + 3H^+]^{+3}$, *22*. $[B3\text{-}CRP32 + Na^+ + 2H^+]^{+3}$, *23*. $[B3\text{-}CRP32 + K^+ + 2H^+]^{+3}$, *24.* $[B3\text{-}CRP32 + Na^+ + K^+ + H^+]^{+3}$, *25*. $[B3\text{-}CRP32 + 2Na^+ + K^+]^{+3}$, *26*. $[2B3\text{-}CRP32 - NH_4^+ + 4H^+]^{+3}$, *27*. $[2B3\text{-}CRP32 + 3H]^{+3}$, *28*. $[2B3\text{-}CRP32 + Na^+ + 2H^+]^{+3}$, *29*. $[2B3\text{-}CRP32 + K^+ + 2H^+]^{+3}$, *30.* $[B3\text{-}CRP32 + Na^+ + K^+ + H^+]^{+3}$, *31*. $[2B3\text{-}CRP32 + 2Na^+ + K^+]^{+3}$, *32*. $[4B3\text{-}CRP32 + 4H^+]^{+4}$, *33*. $[4B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *34*. $[4B3\text{-}CRP32 + Na^+ + K^+ + 2H^+]^{+4}$, *35*. $[4B3\text{-}CRP32 + 2K^+ + 2H^+]^{+4}$, *36*. $[3B3\text{-}CRP32 + 4H^+]^{+4}$, *37*. $[3B3\text{-}CRP32 + K^+ + 3H^+]^{+4}$, *B3* - Procyanidin dimer B3, *CRP32* - QLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF.

The same procedure was applied to determine the identity of peptide-tannin non-covalent complexes for each collected peptidic fraction in the presence of procyanidin B3. When procyanidin B3 was added, some differences were found between the original and subsequent (mixture) ESI-MS spectra. Several new m/z peaks appeared in

those mixtures that are supposed to be the result of B3-peptide fractions complex formation. An example of the differences found between the mass spectra profile of a peptide fraction and their complexes with procyanidin B3 is shown in *Fig. 27*. The ESI-MS mass profile totally changed between control (*Pep Mix6*) and sample (B3-*Pep Mix6* mixture), although the major m/z peaks were kept after B3 addition. Those, in most of the cases, experienced a decrease reinforcing the involvement of *Pep Mix6* peptides in the formation of the new detected analytes. Several new molecular masses appeared in the B3-*Pep Mix6* mixture assay and they were supposed to represent the formed complexes. It should be noted that the new analytes displayed in *Fig. 27* and extended to the other peptide fractions appeared multi-charged. For example, peak 3 (1364 m/z) corresponds to a double charged analyte, so the molecular weight resulting on its deconvolution is 2727 Da (*Table 11*). As shown in *Fig. 27*, the major non-covalent complexes found were the displayed with numbers of 1 to 14. These complexes appeared well differentiated in the B3-*Pep Mix6* spectra being the major (higher ESI signal) the complex codified as 3 followed by 2, 4, 11, 13 and 12 respectively. According to *Table 8*, all these major complexes involved, or may involve, peptides containing CD epitopes, with the exception of the coded as 2. Although the 14-coded complex appeared to be quantitatively relevant in mass spectra it was not characterized.

The same behavior was observed in the other B3-peptide fraction mixtures assays. *Table 11* shows the comprehensive list of the identities of the complexes formed between the sequenced gluten-derived peptides and procyanidin B3 for all collected peptide fractions. The real molecular weights resulting from the deconvolution of the spectra molecular masses is also displayed in *Table 11* along with their hypothesized identity. As represented in *Fig. 28*, the number of identified complexes increased from *Pep Mix1* to *Pep Mix7*. Thus, when *Pep Mix1* and *Pep Mix2* mixture assays were analyzed, only 12% and 9% of their complexes had its composition determined. From *Pep Mix3* to *7* mixture assays about 45%, 73%, 61%, 80% and 100% of total complexes were respectively identified. As aforementioned, only the last four fractions had potential relevance on a CD point of view because of their content in peptides possessing previously described CD epitopes. Interestingly, the number of complexes involving peptides with CD epitopes also increased in the last B3-peptide mixture assays (13%, 32%, 38% and 100% from *Pep Mix4* to *7*, respectively). Therefore, the potential importance of the B3-peptide mixture assays from an immunologic point of view was found in *Pep Mix4* up to *Pep Mix7* (*Fig. 28*).

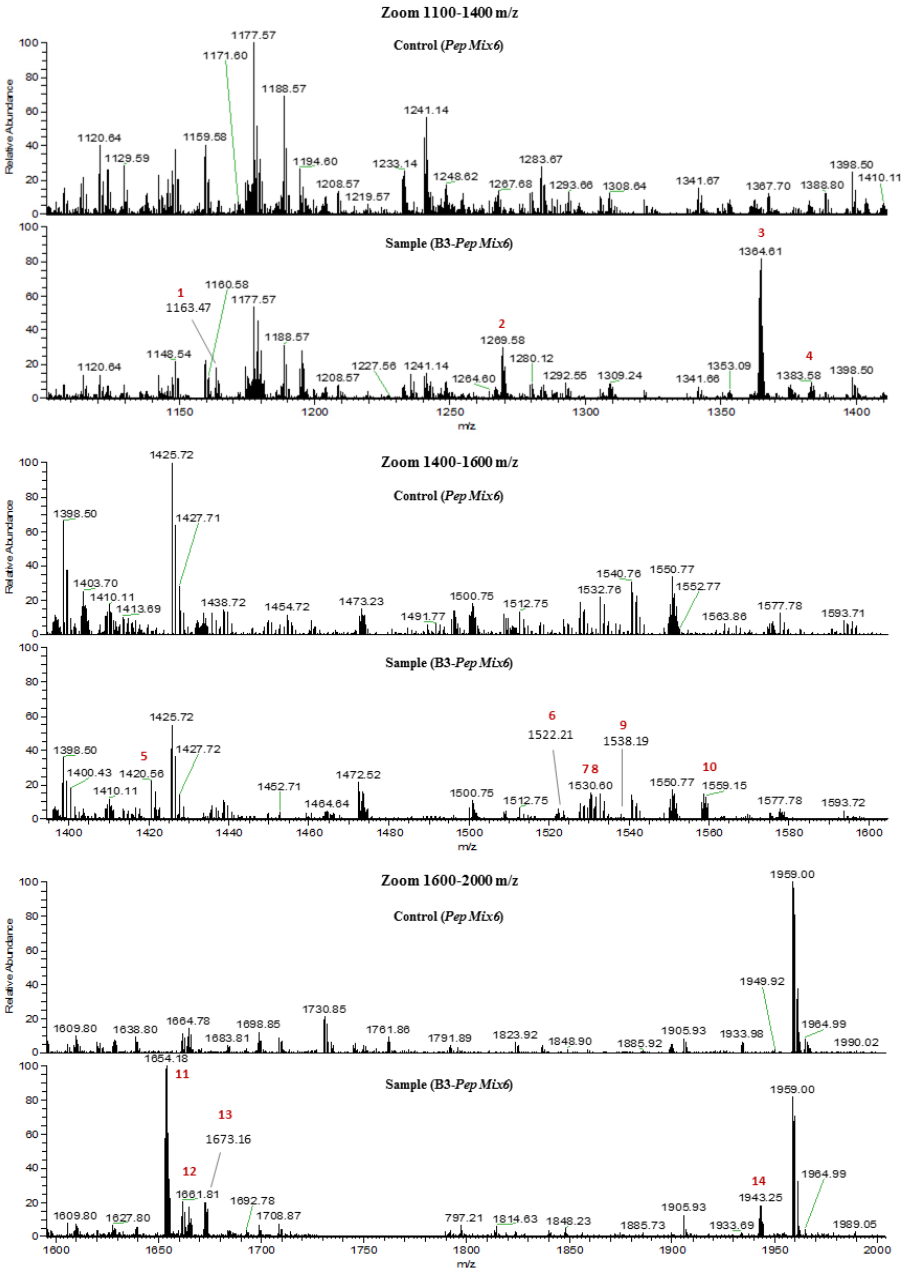


Fig. 27 – ESI-MS spectra of *Pep Mix6* and *B3-Pep Mix6* mixture assay. The most pronounced complexes (only 14 out of 20) are highlighted by red numbers and further characterized in terms of their composition. *1.* $[DPLGAL + B3]^{+1}$, *2.* $[LQPQNPSQQQPQEQVPL + B3 + 2H^+]^{+2}$, *3.* $[QQSSYQVL + 3B3 + K^+ + H^+]^{+2}$ *or* $[QLEMMTSI + 3B3 + K^+ + H^+]^{+2}$ *or* $[SSSPQQLGQGQPRY + 2B3 + K^+ + H^+]^{+2}$ *or* **$[TQQPQQPFPQQPQQPFPQ + B3 + 2H^+]^{+2}$** *or* **$[MQLQPFPQPQLPYPQPQL + B3 + 2H^+]^{+2}$**, *4.* $[QQSSYQVL + 3B3 + 2K^+]^{+2}$ *or* $[QLEMMTSI + 3B3 + 2K^+]^{+2}$ *or* $[SSSPQQLGQGQPRY + 2B3 + 2K^+]^{+2}$ *or* **$[TQQPQQPFPQQPQQPFPQ + B3 + K^+ + H^+]^{+2}$** *or* **$[MQLQPFPQPQLPYPQPQL + B3 + K^+ + H^+]^{+2}$**, *5.* $[GQVQWPQ + B3 + H^+]^{+1}$, *6.*Unknown, *7.* Unknown, *8.* $[QQSSYQVL + B3 + H^+]+1$ *or* $[QLEMMTSI + B3 + H^+]^{+1}$, *9.* Unknown, *10.* $[LQPQNPSQQQPQEQVPL + 2B3 + Na^+ + H^+]^{+2}$, *11.* $[SSSPQQLGQGQPRY + 3B3 + K^+ + H^+]^{+2}$ *or* **$[TQQPQQPFPQQPQQPFPQ + 2B3 + 2H^+]^{+2}$** *or* **$[MQLQPFPQPQLPYPQPQL + 2B3 + 2H^+]^{+2}$**, *12.* **$[TQQPQQPFPQQPQQPFPQ + 2B3 + Na^+ + H^+]^{+2}$**, *13.* $[SSSPQQLGQGQPRY + 3B3 + 2K^+]^{+2}$ *or* **$[TQQPQQPFPQQPQQPFPQ + 2B3 + K^+ + H^+]^{+2}$** *or* **$[MQLQPFPQPQLPYPQPQL + 2B3 + K^+ + H^+]^{+2}$**, *14.* Unknown.

Table 11 – Identification of B3-peptide non-covalent complexes by ESI-MS. The molecular weight (Da) as well as their hypothesized identification is shown. The complexes involving peptides with CD epitopes are highlighted in red.

Pep Mix1	[**1164.41Da**] GQASPQ + B3 or TLSASQ + B3 - H3O$^+$; [**1197.41Da**] AGQGPQ + B3 + Na$^+$ + K$^+$ or RPGCST + B3; [**1241.43Da**] GQASPQ + B3 + 2K$^+$ or MCSVSV + B3 + K$^+$; [**1287.38Da**] QQQGLG + B3 + 2K$^+$; [**1671.56Da**] GNLGGV + 2B3 or QSRQGQQPGQ + B3 - H_3O^+ and [**2109.74Da**] RPSQQNPQ + 2B3
Pep Mix2	[**1308.23Da**] HVSVEH + B3 + Na$^+$; [**1360.55Da**] PQLQNPS + B3; [**1415.59Da**] SQQQQPV + B3 + Na$^+$ or AQGSVQPQ + B3 + Na$^+$; [**1714.57Da**] SPVAGQ + 2B3; [**1770.60Da**] GSVQPQ + 2B3 or AANMQV + 2B3 - H_3O^+ or AANMQV + 2B3 - NH_4^+; [**1953.57Da**] SQQQQPV + 2B3 - NH_4^+or AQGSVQPQ + 2B3 - NH_4^+ or LQPKNPSQQQPQ + B3 - NH_4^+; [**2182.82Da**] LSGQGQRPGQ + 2B3; [**2244.82Da**] LSGQGQRPGQ + 2B3 + Na$^+$ + K$^+$ or SHHQQQQPV + 2B3
Pep Mix3	[**1107.42Da**] VGIGGQ + B3 or DVGTGV + B3 - NH_4^+; [**1142.40Da**] PPDTGG + B3 + Na$^+$ or TYPGGA + B3 or ALGRAP + B3 - H_3O^+ or ALGRAP + B3 - NH_4^+;[**1180.44Da**] TYPGGA + B3 + K$^+$ or GVGTGVGG + B3; [**1182.45Da**] PPDTGG + B3 + Na$^+$ + K$^+$ or TYPGGA + B3 + K$^+$; [**1193.44Da**] AAQLPA + B3 + 2Na$^+$; [**1199.42Da**] PPDTGG + B3 + 2K$^+$ or ALGRAP + B3 + K$^+$; [**1269.46Da**] SIGTGVGG + B3 + 2Na$^+$ or YPTSPQ + B3; [**1367.48Da**] IPSVATY + B3 + K$^+$ or QDQQSGQ + B3 or NMQVDPS + B3; [**1374.55Da**] IPSVATY + B3 + 2Na$^+$ or SQPQQPI + B3; [**1432.52Da**] YYPTSPQ + B3; [**1848.66Da**] SIGTGVGG + 2B3 + 2Na$^+$ or YPTSPQ + 2B3; [**2031.86Da**] SQPQQPI + 2B3 + 2K$^+$; [**2041.88Da**] SQPQHPI + 2B3 + 2K$^+$ or QPRQPFLQPQQP + B3; [**2049.84Da**] YYPTSPQ + 2B3 + K$^+$; [**2138.76Da**] QTFPHQPQ + 2B3; [**2149.80Da**] QPRQPYPQ + 2B3 - H_3O^+ or SQQPGQGQQGYYSGS + B3; [**2169.80Da**] QPRQPYPQ + 2B3; [**2173.90Da**] SQQPGQGQQGYYSGS + B3 + Na$^+$; [**2188.86Da**] SQQPGQGQQGYYSGS + B3 + K+; [**2372.98Da**] MNIQVDPSSQV + 2B3or GSFQPSQQNPQ + 2B3 and [**2620.02Da**] QPRQPFLQPQQP + 2B3
Pep Mix4	[**1232.51Da**] GIIQPQ + B3; [**1241.47Da**] VSFQPS + B3 or QQQPPS + B3 - H_3O^+; [**1272.04Da**] GIIQPQ + B3 + K$^+$; [**1281.50Da**] VSFQPS + B3 + K$^+$ or YPISPQ + B3; [**1283.44Da**] QQQPPS + B3 + Na$^+$; [**1398.55Da**] QPSHQQP + B3; [**1420.53Da**] QPSHQQP + B3 + Na$^+$; [**1528.66Da**] SHQQQPF + B3 + 2K+ or GIIQPQQPA + B3; [**1542.58Da**] QSRCNVMQ + B3 or QQPSYSQQ + B3; [**1980.80Da**] VQGQGIIQPQQPA + B3 + K+ or PQQYSPYQPQQ + B3 + K+; [**2001.84Da**] QPSHQQP + 2B3 + Na$^+$ or QPIPQPQQPQQP + B3 + K$^+$ or **TQQPQQPFPQQP + B3**; [**2009.80Da**] SHQQQPF + 2B3 - NH_4^+ or SHQQQPF + 2B3 - H_3O^+ or QPIPQPQQPQQP + B3 + 2Na$^+$; [**2013.80Da**] VRVPVPQ + 2B3 + Na$^+$ + K$^+$; [**2042.88Da**] QPIPQPQQPQQP + B3 + 2K$^+$ or SQQQQPVLPQQPS + B3 + Na$^+$; [**2064.86Da**] SKQPQQPFPQPQ + B3 + 2K$^+$ or **TQQPQQPFPQQP + B3 + Na$^+$ + K$^+$** or SQQQQPVLPQQPS + B3 + Na$^+$; [**2171.90Da**] LQQQQFPQ + 2B3 or CSPTPYVQSQMWQ + B3 + K$^+$ or QQGQQPGQGQQGQQP + B3; [**2191.90Da**] LLQQCRPV + 2B3 + 2K$^+$; [**2340.98Da**] **SQQPQQPFPQPQQPQ + B3**; [**2375.12Da**] SQQQPPFSQQ + 2B3 + 2Na$^+$ or QQGYDNPYHV + 2B3 or EQQPGQGQEGY + 2B3 or ECCSRLGQMPP + 2B3; [2558.96Da] VQGQGIIQPQQPA + 2B3 + K$^+$ or PQQYSPYQPQQ + 2B3 + K$^+$; [**2587.94Da**] QPIPQPQQPQQP + 2B3 + 2Na$^+$ or SKQPQQPFPQPQ + 2B3 + Na$^+$; [**2643.00Da**] SKQPQQPFPQPQ + 2B3 + 2K$^+$ or SQQQQPVLPQQPS + 2B3 + Na$^+$; [**2795.10Da**] QQGQQPGQGQQGQQP + 2B3 + 2Na$^+$ or SRQQPLPPQQTLSH + 2B3 + Na$^+$ or SHIPGLERPSQQQPLPPQQT + B3 - H_3O^+ and [**2854.02Da**] SHIPGLERPSQQQPLPPQQT + B3 + K$^+$
Pep Mix5	[**1280.48Da**] SSCQSM + B3 + Na$^+$ + K$^+$ or SQQPPF + B3; [**1388.60Da**] VQQQLPV + B3; [**1400.46Da**] NPQAQGSV + B3 + Na$^+$ or PQLQQPL + B3; [**1408.54Da**] TTTSVPF + B3 + 2K$^+$; [**1409.38Da**] TTTSVPF + B3 + 2K$^+$; [**1473.48Da**] PEQEQQP + B3 + K$^+$; [**1528.66Da**] SQQPPFPQ + B3 + Na$^+$ or GIIQPQQPA + B3; [**2013.86Da**] VQQQLPV + 2B3 + 2Na$^+$ or **LQQPQQPFPQPQ + B3** or LQPGQGQQGYYPT + B3; [**2052.90Da**] **LQQPQQPFPQPQ + B3 + K$^+$** or LQPGQGQQGYYPT + B3 + K$^+$ or SQQQQPVLPQQPP + B3; [**2066.84Da**] SQQPPFPQ + 2B3 - H_3O^+ or SQQPPFPQ + 2B3 - NH_4^+ or **SQQPQQPFPQQP + B3 + 2K$^+$** or YQQPQQTFPQPQ + B3; [**2145.94Da**] SQQPPFPQ + 2B3 + Na$^+$ + K$^+$ or QEQQGQQP + 2B3 + 2Na$^+$ or GIIQPQQPA + 2B3 + K$^+$ or FSQQQQTV + 2B3 + Na$^+$ or HSKQGQQPG + 2B3 + Na$^+$ or YQQPQQTFPQPQ + B3 + 2K$^+$; [**2184.78Da**] GIIQPQQPA + 2B3 + 2K$^+$ or FSQQQQTV + 2B3 + Na$^+$ + K$^+$ or HSKQGQQPG + 2B3 + Na$^+$ + K$^+$ or FQQPHQPF + 2B3 or ILQQQQQQQQQQV + B3 - H_3O^+ or ILQQQQQQQQQQV + B3 - NH_4^+; [**2452.10Da**] AATPNSSTEVMTS + 2B3 or AQMETNCISGLE + 2B3 or QQQPGQGQQGHC + 2B3 or QPEQGQEGYYP + 2B3; [**2537.14Da**] LQPQNPSQQQPQEQVPL + B3; [**2560.14Da**] **TQQPQQPFPQQP + 2B3 - H3O$^+$** or **TQQPQQPFPQQP + 2B3 - NH_4^+** or LQPQNPSQQQPQEQVPL + B3 + Na$^+$; [**2592.02Da**] **LQQPQQPFPQPQ + 2B3** or LQPGQGQQGYYPT + 2B3; [**2617.94Da**] **TQQPQQPFPQQP + 2B3 + K$^+$** or QSISQYQQQQPQ + 2B3; [**2762.92Da**] ILQQQQQQQQQQV + 2B3 - H_3O^+ or ILQQQQQQQQQQV + 2B3 - NH4+ and [**2816.29Da**] SHIPGLERPSQQQPLPPQQT + B3
Pep Mix6	[**1162.46Da**] DPLGAL + B3; [**1291.54Da**] ALETLPA + B3; [**1376.52Da**] QPSQQNP + B3; [**1419.55Da**] GQVQWPQ + B3; [**1434.59Da**] QIPEQSR + B3; [**1436.36Da**] QPSQQNP + B3 + Na$^+$ + K$^+$; [**1529.59Da**] QQSSYQVL + B3 or QLEMMTSI + B3; [**2537.14Da**] LQPQNPSQQQPQEQVPL + B3; [**2727.20Da**] QQSSYQVL + 3B3 + K$^+$ or QLEMMTSI + 3B3 + K$^+$ or SSSPQQLGQGQPRY + 2B3 + K$^+$ or **TQQPQQPFPQQPQQPFPQ + B3** or **MQLQPFPQPQLPYPQPQL + B3**; [**2749.18Da**] QQSSYQVL + 3B3 + Na$^+$ + K$^+$ or QLEMMTSI + 3B3 + Na$^+$ + K$^+$ or SSSPQQLGQGQPRY + 2B3 + Na$^+$ + K$^+$ or **TQQPQQPFPQQPQQPFPQ + B3 + Na$^+$** or **MQLQPFPQPQLPYPQPQL + B3 + Na$^+$**; [**2765.14Da**] QQSSYQVL + 3B3 + 2K$^+$ or QLEMMTSI + 3B3 + 2K$^+$ or SSSPQQLGQGQPRY + 2B3 + 2K$^+$ or **TQQPQQPFPQQPQQPFPQ + B3 + K$^+$** or **MQLQPFPQPQLPYPQPQL + B3 + K$^+$**; [**3137.28Da**] LQPQNPSQQQPQEQVPL + 2B3 + Na$^+$; [**3154.24Da**] LQPQNPSQQQPQEQVPL + 2B3 + K$^+$; [**3306.34Da**] SSSPQQLGQGQPRY + 3B3 + K$^+$ or **TQQPQQPFPQQPQQPFPQ + 2B3** or **MQLQPFPQPQLPYPQPQL + 2B3**; [**3327.32Da**] **TQQPQQPFPQQPQQPFPQ + 2B3 + Na$^+$**;[**3344.30Da**] SSSPQQLGQGQPRY + 3B3 + 2K$^+$ or **TQQPQQPFPQQPQQPFPQ + 2B3 + K$^+$** or **MQLQPFPQPQLPYPQPQL + 2B3 + K$^+$**
Pep Mix7	**[3795Da]** **TQQPQQPFPQQPQQPFPQQPQQPFPQ +B3 + K$^+$**

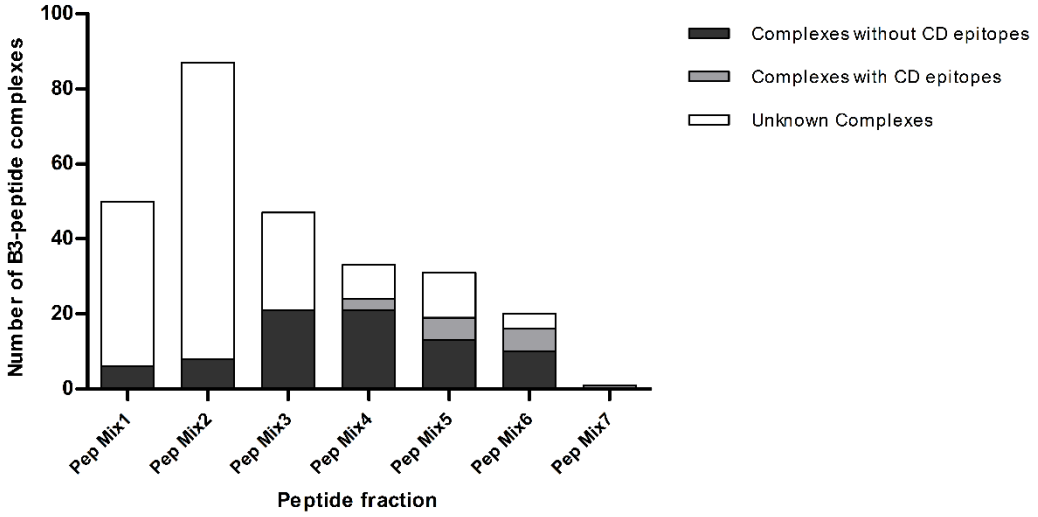


Fig. 28 – Qualitative parameters related to the identification of B3-gliadin derived peptide complexes: number of unknown and identified non-covalent complexes (with or without epitopes).

If we focus in the complexes formed between procyanidin B3 and peptides containing CD T cell epitopes, we could establish several differences among mixtures (*Table 11*). Thus, in *B3-Pep Mix4* mixture assay there were two different peptides involved in complex formation (two complexes containing TQQPQQPFPQQP and one SQQPQQPFPQPQQPQ). Both had one copy of the *a-type* epitope. *B3-Pep Mix5* mixture assay presented six different complexes: two with TQQPQQPFPQQP, one with SQQPQQPFPQQP and two other with LQQPQQPFPQPQ. All contained one copy of the *a-type* CD epitope. In *B3-Pep Mix6* mixture assay, 6 distinct complexes were identified. Two different peptides were hypothesized as being part of five of the six complexes: TQQPQQPFPQQPQQPFPQ (one copy of the *a*- and *b-type* epitopes) and MQLQPFPQPQLPYPQPQL (one copy of the *c*- and *d-type* epitopes). In this mixture, the TQQPQQPFPQQPQQPFPQ peptide was additionally present in a complex with a molecular weight of 3137.28 Da. For *B3-Pep Mix7* mixture assay, only the TQQPQQPFPQQPQQPFPQQPQQPFPQ + B3 (+ K^+) complex was detected. The peptide involved in this complex contains two copies of *a-type* and one copy of b-*type* epitope. The number of procyanidin B3 units involved in complex formation varied from one to three moieties. Interestingly, all the peptides containing epitopes from the last four fractions that were involved in complex formation were identified as gamma-gliadin derivatives regardless of the fraction tested (except the MQLQPFPQPQLPYPQPQL peptide, in *Pep Mix6* assay, that belongs to an alpha-gliadin) (*Table 8*). This appears to indicate that gamma-gliadins are the most representative protein subtype in the crude

extract. In fact, gamma-gliadins peptides appeared homogenously distributed along the different collected peptidic fractions (*Fig. 20*).

Overall, the majority of complexes between procyanidin B3 and peptides containing epitopes were found in *Pep Mix5* and *6* (six complexes per peptide fraction). Although in *Pep Mix7* only one complex has been detected by ESI-MS, it involved a multivalent peptide with 3 CD epitopes. This kind of long and oligomerized peptides are supposed to be very effective in T cell stimulation as they provide multivalent TG2 substrates [16, 28].

Among composition, in general, differences were also observed in terms of the complexes molecular weight and number. According to *Fig. 29*, from *Pep Mix1* to *Pep Mix7* there was an overall decrease in the number of detected complexes while their size became gradually higher. Thus, the number of new complexes in *Pep Mix2* assay was near to ninety while in *Pep Mix7* only one non-covalent complex was found (highest one).

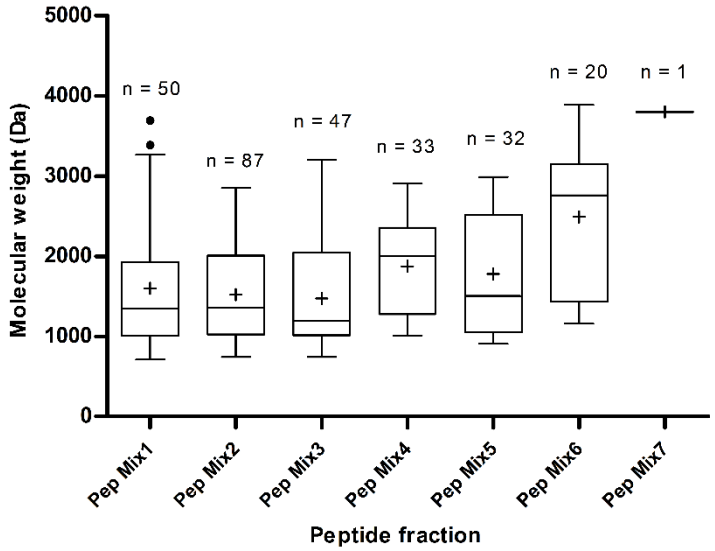


Fig. 29 – Molecular weight (Da) of B3-gliadin derived peptide complexes found in the different peptidic fraction assays. The average molecular weight of each fraction non-covalent complexes is displayed by the + symbol.

As above mentioned, the ability of tannins to complex with the peptide fractions was verified by dynamic light scattering experiments (*Fig. 20*). The complexes size showed therein experienced an appreciable increase in *Pep Mix6* assay reaching a maximum size value in *Pep Mix7* test. During the *Pep Mix7* assay, the instrument lecture disrupted between replicate measurements because those mixtures experienced a faster precipitation process and consequently this fraction had to be discarded for analysis. Those results may be in agreement with the ones showed in here since ESI-MS is only probed as an optimal analytical method for the characterization and

identification of soluble tannin-protein complexes [120, 121, 213]. Therefore, the results presented herein confirms, for the first time, the identity of the soluble non-covalent complexes between the tannin procyanidin B3 and different gluten-derived peptides but did not highlight the identity of the higher insoluble ones that may occur in B3-*Pep Mix7* mixture assay.

In summary, several CD relevant soluble complexes were identified in the last four B3-peptide mixtures assays from both a qualitative and quantitative point of view. Keeping in mind that the peptides containing CD epitopes represented a slight percentage of those available to interaction, the observation that they were involved in the formation of a high number of non-covalent complexes highlights a binding affinity of procyanidin B3 to complex with peptides containing CD epitopes.

CONCLUSIONS

Celiac disease (CD) is an immune-mediated disorder induced by the ingestion of gluten in genetically susceptible people. The increasing number of individuals being diagnosed with some form of sensitivity to wheat cereal grain proteins represents a cause for concern. Currently, the only accepted treatment for CD involves the dietary withdrawal of gluten; however, the response to therapy is poor in up to 30% of patients, and dietary nonadherence is the chief cause of persistent or recurrent symptoms. Worthy a special attention, these cases require extensive evaluation to rule out intestinal lymphoma and refractory sprue, complications that arise as the result of clonal expansion of intraepithelial lymphocytes. The high content of glutamine is necessary for the plant as a nitrogen source, but the massive proportion of proline residues makes gluten proteins quite resistant to digestion by mammalian proteases. In fact, this amino acid has a cyclic side chain, whose steric hindrance limits the access of proteolytic enzymes to the peptidic bonds. This leads to the acumulation of medium to high molecular weight peptides in the gastrointestinal tract, some of which are known to be involved in triggering celiac disease.

Tannins are astringent, bitter plant polyphenolic compounds that binds to and precipitates proteins and various other organic compounds including amino acids and alkaloids. Given that tannins are specially reactive towards proline-rich proteins and considering that the CD-related peptides have high proline contents, the main (and global) goal of this work was to study the interaction between tannins and those immunotoxic peptides.

Altogether, it was found that:

1. Dietary tannins are able to complex with different wheat-derived peptides, as verified using two different techniques to cover a large range of concentrations: from the micromolar range with fluorescence to the millimolar range with DLS;
2. That aggregation (or non-covalent complexation) depends on the tannin tested and varied along with the tannin and protein concentration;
3. The immunoreactive peptides identified after *in vitro* gluten digestion revealed a high affinity to bind with procyanidin B3. Nevertheless, a much deeper approach will be necessary to differentiate those products that are indeed important from a disease point of view and also to specifically study their higher or lower propensity to react with food tannins.

Overall, this study provides an initial insight about the potential role of phenolic compounds as a nutritional therapy for CD, thus generating new perspectives about its

applicability as modulators of this chronic inflammatory condition. The next steps will require further biological studies involving these peptides in the presence of different polyphenols to assess the physiological and biochemical consequences of the association process described herein in terms of how it does interfere with those peptides ability/availability to damage the intestinal mucosa.

REFERENCES

1. Arentz-Hansen, H., et al., *The molecular basis for oat intolerance in patients with Celiac disease.* Plos Medicine, 2004. **1**(1): p. 84-92.
2. Sollid, L.M. and C. Khosla, *Future therapeutic options for celiac disease.* Nature Clinical Practice Gastroenterology & Hepatology, 2005. **2**(3): p. 140-147.
3. Fasano, A. and C. Catassi, *Current approaches to diagnosis and treatment of celiac disease: An evolving spectrum.* Gastroenterology, 2001. **120**(3): p. 636-651.
4. Ferretti, G., et al., *Celiac Disease, Inflammation and Oxidative Damage: A Nutrigenetic Approach.* Nutrients, 2012. **4**(4): p. 243-257.
5. Green, P.H.R. and C. Cellier, *Celiac Disease.* New England Journal of Medicine, 2007. **357**(17): p. 1731-1743.
6. Sollid, L.M., *Molecular basis of celiac disease.* Annual Review of Immunology, 2000. **18**: p. 53-81.
7. Calder, P.C., et al., *Inflammatory Disease Processes and Interactions with Nutrition.* British Journal of Nutrition, 2009. **101**: p. S1-+.
8. Corrao, G., et al., *Mortality in patients with coeliac disease and their relatives: a cohort study.* Lancet, 2001. **358**(9279): p. 356-361.
9. Maki, M. and P. Collin, *Coeliac disease.* Lancet, 1997. **349**(9067): p. 1755-1759.
10. Shan, L., et al., *Structural basis for gluten intolerance in Celiac sprue.* Science, 2002. **297**(5590): p. 2275-2279.
11. Rubio-Tapia, A., et al., *ACG Clinical Guidelines: Diagnosis and Management of Celiac Disease.* American Journal of Gastroenterology, 2013. **108**(5): p. 656-676.
12. Bernardo, D. and A.S. Pena, *Developing strategies to improve the quality of life of patients with gluten intolerance in patients with and without coeliac disease.* European Journal of Internal Medicine, 2012. **23**(1): p. 6-8.
13. Gujral, N., H.J. Freeman, and A.B.R. Thomson, *Celiac disease: Prevalence, diagnosis, pathogenesis and treatment.* World Journal of Gastroenterology, 2012. **18**(42): p. 6036-6059.
14. Rewers, M., *Epidemiology of celiac disease: What are the prevalence, incidence, and progression of celiac disease?* Gastroenterology, 2005. **128**(4): p. S47-S51.
15. van Heel, D.A. and J. West, *Recent advances in coeliac disease.* Gut, 2006. **55**(7): p. 1037-1046.

16. Sollid, L.M., *Coeliac disease: Dissecting a complex inflammatory disorder.* Nature Reviews Immunology, 2002. **2**(9): p. 647-655.
17. Shewry, P.R., et al., *The classification and nomenclature of wheat gluten proteins - a reassessment.* Journal of Cereal Science, 1986. **4**(2): p. 97-106.
18. Koehler, P. and H. Wieser, *Chemistry of Cereal Grains.* 2013: p. 11-45.
19. Shewry, P.R. and N.G. Halford, *Cereal seed storage proteins: structures, properties and role in grain utilization.* Journal of Experimental Botany, 2002. **53**(370): p. 947-958.
20. Wang, J.-S., et al., *Characterization of hydrolysates derived from enzymatic hydrolysis of wheat gluten.* Journal of Food Science, 2007. **72**(2): p. C103-C107.
21. Day, L., et al., *Wheat-gluten uses and industry needs.* Trends in Food Science & Technology, 2006. **17**(2): p. 82-90.
22. *The vegetable proteins. By Thomas B. Osborne. Pp. xiii+154. 2nd edition. Monographs on Biochemistry. London: Longmans green and Co. 1924. Pric 9s. net.* Journal of the Society of Chemical Industry, 1924. **43**(17): p. 440-440.
23. Caminero, A., et al., *Chapter 13 - Gluten Metabolism in Humans: Involvement of the Gut Microbiota*, in *Wheat and Rice in Disease Prevention and Health*, R.R.W.R.P. Zibadi, Editor. 2014, Academic Press: San Diego. p. 157-170.
24. Wieser, H., *Chemistry of gluten proteins.* Food Microbiology, 2007. **24**(2): p. 115-119.
25. Ferranti, P., et al., *Mass spectrometry analysis of gliadins in celiac disease.* Journal of Mass Spectrometry, 2007. **42**(12): p. 1531-1548.
26. Payne, P.I., *Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality.* Annual Review of Plant Physiology and Plant Molecular Biology, 1987. **38**: p. 141-153.
27. Wieser, H., S. Antes, and W. Seilmeier, *Quantitative determination of gluten protein types in wheat flour by reversed-phase high-performance liquid chromatography.* Cereal Chemistry, 1998. **75**(5): p. 644-650.
28. Shan, L., et al., *Identification and analysis of multivalent proteolytically resistant peptides from gluten: Implications for Celiac Sprue.* Journal of Proteome Research, 2005. **4**(5): p. 1732-1741.
29. Wieser, H., *Relation between gliadin structure and coeliac toxicity.* Acta paediatrica (Oslo, Norway : 1992). Supplement, 1996. **412**: p. 3-9.
30. Banc, A., et al., *Structure and Orientation Changes of ω- and γ-Gliadins at the Air−Water Interface: A PM−IRRAS Spectroscopy and Brewster Angle Microscopy Study.* Langmuir, 2007. **23**(26): p. 13066-13075.

31. Grosch, W. and H. Wieser, *Redox reactions in wheat dough as affected by ascorbic acid.* Journal of Cereal Science, 1999. **29**(1): p. 1-16.
32. Tatham, A.S. and P.R. Shewry, *The conformation of wheat gluten proteins - the secondary structures and thermal stabilities of alpha-gliadins,beta-gliadins,gamma-gliadins and omega-gliadins.* Journal of Cereal Science, 1985. **3**(2): p. 103-113.
33. Wieser, H. and R. Kieffer, *Correlations of the amount of gluten protein types to the technological properties of wheat flours determined on a micro-scale.* Journal of Cereal Science, 2001. **34**(1): p. 19-27.
34. Wrigley, C.W., *Biopolymers - Giant proteins with flour power.* Nature, 1996. **381**(6585): p. 738-739.
35. Wieser, H., W. Bushuk, and F. MacRitchie, *Chapter 7 The Polymeric Glutenins*, in *Gliadin and Glutenin: The Unique Balance of Wheat Quality*. 2006, AACC International, Inc. p. 213-240.
36. Wieser, H., *18 - The use of redox agents in breadmaking*, in *Breadmaking (Second edition)*, S.P. Cauvain, Editor. 2012, Woodhead Publishing. p. 447-469.
37. Veraverbeke, W.S. and J.A. Delcour, *Wheat protein composition and properties of wheat glutenin in relation to breadmaking functionality.* Critical Reviews in Food Science and Nutrition, 2002. **42**(3): p. 179-208.
38. Shewry, P.R., N.G. Halford, and A.S. Tatham, *High-molecular-weight subunits of wheat glutenin.* Journal of Cereal Science, 1992. **15**(2): p. 105-120.
39. Howdle, P.D., *Gliadin, glutenin or both? The search for the Holy Grail in coeliac disease.* European Journal of Gastroenterology & Hepatology, 2006. **18**(7): p. 703-706.
40. Shewry, P.R. and A.S. Tatham, *The prolamin storage proteins of cereal seeds - structure and evolution.* Biochemical Journal, 1990. **267**(1): p. 1-12.
41. Hausch, F., et al., *Intestinal digestive resistance of immunodominant gliadin peptides.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2002. **283**(4): p. G996-G1003.
42. Fanciulli, G., et al., *Liquid chromatography-mass spectrometry assay for quantification of Gluten Exorphin B5 in cerebrospinal fluid.* Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences, 2007. **852**(1-2): p. 485-490.
43. Fukudome, S., et al., *Release of opioid peptides, gluten exorphins by the action of pancreatic elastase.* Febs Letters, 1997. **412**(3): p. 475-479.

44. Green, P.H.R. and C. Cellier, *Medical progress: Celiac disease.* New England Journal of Medicine, 2007. **357**(17): p. 1731-1743.
45. Sollid, L.M. and B.A. Lie, *Celiac disease genetics: Current concepts and practical applications.* Clinical Gastroenterology and Hepatology, 2005. **3**(9): p. 843-851.
46. Heap, G.A. and D.A. van Heel, *Genetics and pathogenesis of coeliac disease.* Seminars in Immunology, 2009. **21**(6): p. 346-354.
47. Ivarsson, A., *The Swedish epidemic of coeliac disease explored using an epidemiological approach - some lessons to be learnt.* Best Practice & Research in Clinical Gastroenterology, 2005. **19**(3): p. 425-440.
48. Stene, L.C., et al., *Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: A longitudinal study.* American Journal of Gastroenterology, 2006. **101**(10): p. 2333-2340.
49. Di Sabatino, A. and G.R. Corazza, *Coeliac disease.* Lancet, 2009. **373**(9673): p. 1480-1493.
50. Persson, L.A., A. Ivarsson, and O. Hernell, *Breast-feeding protects against Celiac disease in childhood - Epidemiological evidence*, in *Integrating Population Outcomes, Biological Mechanisms and Research Methods in the Study of Human Milk and Lactation*, M.K. Davis, et al., Editors. 2002. p. 115-123.
51. Harris, P.J. and L.R. Ferguson, *Chapter 16 - Wheat and Rice Dietary Fiber in Colorectal Cancer Prevention and the Maintenance of Health*, in *Wheat and Rice in Disease Prevention and Health*, R.R.W.R.P. Zibadi, Editor. 2014, Academic Press: San Diego. p. 201-210.
52. Ivarsson, A., et al., *Breast-feeding protects against celiac disease.* American Journal of Clinical Nutrition, 2002. **75**(5): p. 914-921.
53. Norris, J.M., et al., *Risk of celiac disease autoimmunity and timing of gluten introduction in the diet of infants at increased risk of disease.* Jama-Journal of the American Medical Association, 2005. **293**(19): p. 2343-2351.
54. Jabri, B., D.D. Kasarda, and P.H.R. Green, *Innate and adaptive immunity: the Yin and Yang of celiac disease.* Immunological Reviews, 2005. **206**: p. 219-231.
55. Koning, F., et al., *Pathomechanisms in celiac disease.* Best Practice & Research in Clinical Gastroenterology, 2005. **19**(3): p. 373-387.
56. Molberg, O., et al., *Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease (vol 4, pg 713, 1998).* Nature Medicine, 1998. **4**(8): p. 974-974.

57. Sollid, L.M. and B. Jabri, *Triggers and drivers of autoimmunity: lessons from coeliac disease.* Nature Reviews Immunology, 2013. **13**(4): p. 294-302.
58. Molberg, O., et al., *T cells from celiac disease lesions recognize gliadin epitopes deamidated in situ by endogenous tissue transglutaminase.* European Journal of Immunology, 2001. **31**(5): p. 1317-1323.
59. Klock, C., T.R. DiRaimondo, and C. Khosla, *Role of transglutaminase 2 in celiac disease pathogenesis.* Seminars in Immunopathology, 2012. **34**(4): p. 513-522.
60. Fleckenstein, B., et al., *Gliadin T cell epitope selection by tissue transglutaminase in Celiac disease - Role of enzyme specificity and pH influence on the transamidation versus deamidation reactions.* Journal of Biological Chemistry, 2002. **277**(37): p. 34109-34116.
61. Sollid, L.M. and B. Jabri, *Is celiac disease an autoimmune disorder?* Current Opinion in Immunology, 2005. **17**(6): p. 595-600.
62. Camarca, A., A. Del Mastro, and C. Gianfrani, *Repertoire of gluten peptides active in celiac disease patients: perspectives for translational therapeutic applications.* Endocrine, metabolic & immune disorders drug targets, 2012. **12**(2): p. 207-19.
63. Nilsen, E.M., et al., *Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease.* Gastroenterology, 1998. **115**(3): p. 551-563.
64. Kagnoff, M.F., *Overview and pathogenesis of celiac disease.* Gastroenterology, 2005. **128**(4): p. S10-S18.
65. Hue, S., et al., *A direct role for NKG2D/MICA interaction in villous atrophy during celiac disease.* Immunity, 2004. **21**(3): p. 367-377.
66. Mention, J.J., et al., *Interleukin 15: A key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease.* Gastroenterology, 2003. **125**(3): p. 730-745.
67. Meresse, B., et al., *Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease.* Immunity, 2004. **21**(3): p. 357-366.
68. Osorio, C., et al., *Targeted modification of wheat grain protein to reduce the content of celiac causing epitopes.* Functional & Integrative Genomics, 2012. **12**(3): p. 417-438.
69. Tye-Din, J.A., et al., *Comprehensive, Quantitative Mapping of T Cell Epitopes in Gluten in Celiac Disease.* Science Translational Medicine, 2010. **2**(41).

70. Lee, S.K., et al., *Duodenal histology in patients with celiac disease after treatment with a gluten-free diet.* Gastrointestinal Endoscopy, 2003. **57**(2): p. 187-191.
71. Matysiak-Budnik, T., et al., *Long-term follow-up of 61 coeliac patients diagnosed in childhood: evolution toward latency is possible on a normal diet.* Gut, 2007. **56**(10): p. 1379-1386.
72. Mayer, M., et al., *Compliance of adolescents with Celiac-disease with a gluten free diet.* Gut, 1991. **32**(8): p. 881-885.
73. Bakshi, A., et al., *Emerging therapeutic options for celiac disease: potential alternatives to a gluten-free diet.* Gastroenterology & hepatology, 2012. **8**(9): p. 582-8.
74. Camarca, A., et al., *Intestinal T Cell Responses to Gluten Peptides Are Largely Heterogeneous: Implications for a Peptide-Based Therapy in Celiac Disease.* Journal of Immunology, 2009. **182**(7): p. 4158-4166.
75. Spaenij-Dekking, L., et al., *Natural variation in toxicity of wheat: Potential for selection of nontoxic varieties for celiac disease patients.* Gastroenterology, 2005. **129**(3): p. 797-806.
76. Siegel, M., et al., *Rational design of combination enzyme therapy for celiac sprue.* Chemistry & Biology, 2006. **13**(6): p. 649-658.
77. Stepniak, D., et al., *Highly efficient gluten degradation with a newly identified prolyl endoprotease: implications for celiac disease.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2006. **291**(4): p. G621-G629.
78. Pyle, G.G., et al., *Effect of pretreatment of food gluten with prolyl endopeptidase on gluten-induced malabsorption in celiac sprue.* Clinical Gastroenterology and Hepatology, 2005. **3**(7): p. 687-694.
79. Vanhoof, G., et al., *Cloning and sequence-analysis of the gene encoding human lymphocyte prolyl endopeptidase.* Gene, 1994. **149**(2): p. 363-366.
80. Siegel, M. and C. Khosla, *Transglutaminase 2 inhibitors and their therapeutic role in disease states.* Pharmacology & Therapeutics, 2007. **115**(2): p. 232-245.
81. Aeschlimann, D. and M. Paulsson, *Transglutaminases - protein cross-linking enzymes in tissues and body-fluids.* Thrombosis and Haemostasis, 1994. **71**(4): p. 402-415.
82. Szondy, Z., et al., *Transglutaminase 2(-/-) mice reveal a phagocytosis-associated crosstalk between macrophages and apoptotic cells.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(13): p. 7812-7817.

83. Maurano, F., et al., *Intranasal administration of one alpha gliadin can downregulate the immune response to whole gliadin in mice.* Scandinavian Journal of Immunology, 2001. **53**(3): p. 290-295.
84. Appel, H., et al., *Anergy induction by dimeric TCR ligands.* Journal of Immunology, 2001. **166**(8): p. 5279-5285.
85. Mulder, C.J.J., et al., *A pilot study of recombinant human interieukin-10 in adults with refractory coeliac disease.* European Journal of Gastroenterology & Hepatology, 2001. **13**(10): p. 1183-1188.
86. Salvati, V.M., et al., *Recombinant human interleukin 10 suppresses gliadin dependent T cell activation in ex vivo cultured coeliac intestinal mucosa.* Gut, 2005. **54**(1): p. 46-53.
87. Sette, A., P. Wentworth, and H.M. Grey, *Major histocompatibility complex binding peptides: a target for therapeutic development.* Current opinion in biotechnology, 1991. **2**(6): p. 877-81.
88. Manach, C., et al., *Polyphenols: food sources and bioavailability.* American Journal of Clinical Nutrition, 2004. **79**(5): p. 727-747.
89. de Freitas, V. and N. Mateus, *Protein/Polyphenol Interactions: Past and Present Contributions. Mechanisms of Astringency Perception.* Current Organic Chemistry, 2012. **16**(6): p. 724-746.
90. Lattanzio, V., et al., *Plant Phenolics – Secondary Metabolites with Diverse Functions*, in *Recent Advances in Polyphenol Research*. 2009, Wiley-Blackwell. p. 1-35.
91. Naczk, M. and F. Shahidi, *Phenolics in cereals, fruits and vegetables: Occurrence, extraction and analysis.* Journal of Pharmaceutical and Biomedical Analysis, 2006. **41**(5): p. 1523-1542.
92. Lesschaeve, I. and A.C. Noble, *Polyphenols: factors influencing their sensory properties and their effects on food and beverage preferences.* American Journal of Clinical Nutrition, 2005. **81**(1): p. 330S-335S.
93. Santos-Buelga, C. and A. Scalbert, *Proanthocyanidins and tannin-like compounds - nature, occurrence, dietary intake and effects on nutrition and health.* Journal of the Science of Food and Agriculture, 2000. **80**(7): p. 1094-1117.
94. Joslyn, M.A. and J.L. Goldstein, *Astringency principles - changes in phenolic content in persimmons during ripening + processing.* Journal of Agricultural and Food Chemistry, 1964. **12**(6): p. 511-&.
95. Faria, A., et al., *Procyanidins as antioxidants and tumor cell growth modulators.* Journal of Agricultural and Food Chemistry, 2006. **54**(6): p. 2392-2397.

96. Khafif, A., et al., *Quantitation of chemopreventive synergism between (-)-epigallocatechin-3-gallate and curcumin in normal, premalignant and malignant human oral epithelial cells.* Carcinogenesis, 1998. **19**(3): p. 419-424.
97. Weng, C.-J. and G.-C. Yen, *Chemopreventive effects of dietary phytochemicals against cancer invasion and metastasis: Phenolic acids, monophenol, polyphenol, and their derivatives.* Cancer Treatment Reviews, 2012. **38**(1): p. 76-87.
98. Morley, N., et al., *The green tea polyphenol (-)-epigallocatechin gallate and green tea can protect human cellular DNA from ultraviolet and visible radiation-induced damage.* Photodermatology Photoimmunology & Photomedicine, 2005. **21**(1): p. 15-22.
99. Shim, J.S., et al., *Chemopreventive effect of green tea (camellia-sinensis) among cigarette smokers.* Cancer Epidemiology Biomarkers & Prevention, 1995. **4**(4): p. 387-391.
100. Azevedo, J., et al., *Antioxidant properties of anthocyanidins, anthocyanidin-3-glucosides and respective portisins.* Food Chemistry, 2010. **119**(2): p. 518-523.
101. Wolfe, K.L., et al., *Cellular antioxidant activity of common fruits.* Journal of Agricultural and Food Chemistry, 2008. **56**(18): p. 8418-8426.
102. Akiyama, H., et al., *Antiallergic effect of apple polyphenols on the allergic model mouse.* Biological & Pharmaceutical Bulletin, 2000. **23**(11): p. 1370-1373.
103. Singh, A., S. Holvoet, and A. Mercenier, *Dietary polyphenols in the prevention and treatment of allergic diseases.* Clinical and Experimental Allergy, 2011. **41**(10): p. 1346-1359.
104. Schmitz-Eiberger, M. and A. Matthes, *Effect of harvest maturity, duration of storage and shelf life of apples on the allergen Mal d 1, polyphenoloxidase activity and polyphenol content.* Food Chemistry, 2011. **127**(4): p. 1459-1464.
105. Bellaver, B., et al., *Resveratrol increases antioxidant defenses and decreases proinflammatory cytokines in hippocampal astrocyte cultures from newborn, adult and aged Wistar rats.* Toxicology in Vitro, 2014. **28**(4): p. 479-484.
106. Gu, Y. and J.D. Lambert, *Modulation of metabolic syndrome-related inflammation by cocoa.* Molecular Nutrition & Food Research, 2013. **57**(6): p. 948-961.
107. Bijak, M., M.B. Ponczek, and P. Nowak, *Polyphenol compounds belonging to flavonoids inhibit activity of coagulation factor X.* International Journal of Biological Macromolecules, 2014. **65**: p. 129-135.

108. Carrieri, C., et al., *Antithrombotic activity of 12 table grape varieties. Relationship with polyphenolic profile.* Food Chemistry, 2013. **140**(4): p. 647-653.
109. Zakaria, Z.A., et al., *Methanol Extract of Bauhinia purpurea Leaf Possesses Anti-Ulcer Activity.* Medical Principles and Practice, 2012. **21**(5): p. 476-482.
110. Mulvihill, E.E. and M.W. Huff, *Citrus flavonoids and the prevention of atherosclerosis.* Cardiovascular & hematological disorders drug targets, 2012. **12**(2): p. 84-91.
111. Funatogawa, K., et al., *Antibacterial activity of hydrolyzable tannins derived from medicinal plants against Helicobacter pylori.* Microbiology and Immunology, 2004. **48**(4): p. 251-261.
112. Taguri, T., T. Tanaka, and I. Kouno, *Antimicrobial activity of 10 different plant polyphenols against bacteria causing food-borne disease.* Biological & Pharmaceutical Bulletin, 2004. **27**(12): p. 1965-1969.
113. Joshi, S.S., A.B. Howell, and D.H. D'Souza, *Cronobacter sakazakii reduction by blueberry proanthocyanidins.* Food Microbiology, 2014. **39**: p. 127-131.
114. Pandey, K.B. and S.I. Rizvi, *Plant polyphenols as dietary antioxidants in human health and disease.* Oxidative Medicine and Cellular Longevity, 2009. **2**(5): p. 270-278.
115. Helfer, M., et al., *The Root Extract of the Medicinal Plant Pelargonium sidoides Is a Potent HIV-1 Attachment Inhibitor.* Plos One, 2014. **9**(1).
116. Qiu, X., et al., *Self-Organizing Maps for the Classification of Gallic Acylate Polyphenols as HSV-1 Inhibitors.* Medicinal Chemistry, 2014. **10**(4): p. 388-401.
117. Chung, K.T., et al., *Tannins and human health: A review.* Critical Reviews in Food Science and Nutrition, 1998. **38**(6): p. 421-464.
118. Mueller-Harvey, I., *Unravelling the conundrum of tannins in animal nutrition and health.* Journal of the Science of Food and Agriculture, 2006. **86**(13): p. 2010-2037.
119. McDougall, G.J. and D. Stewart, *The inhibitory effects of berry polyphenols on digestive enzymes.* Biofactors, 2005. **23**(4): p. 189-195.
120. Perez-Gregorio, M.R., N. Mateus, and V. de Freitas, *Rapid Screening and Identification of New Soluble Tannin-Salivary Protein Aggregates in Saliva by Mass Spectrometry (MALDI-TOF-MS-TOF and FIA-ESI-MS).* Langmuir, 2014. **30**(28): p. 8528-8537.
121. Perez-Gregorio, M.R., N. Mateus, and V. De Freitas, *New Procyanidin B3-Human Salivary Protein Complexes by Mass Spectrometry. Effect of Salivary*

Protein Profile, Tannin Concentration, and Time Stability. Journal of Agricultural and Food Chemistry, 2014. **62**(41): p. 10038-10045.

122. de Moura, V.M., et al., *Plants used to treat snakebites in Santarem, western Para, Brazil: An assessment of their effectiveness in inhibiting hemorrhagic activity induced by Bothrops jararaca venom.* Journal of ethnopharmacology, 2015. **161**: p. 224-32.
123. Prabha, B. and N. Savithramma, *Screening of phytochemical constituents of the leaves of Clinacanthus siamensis Bremek and Cissampelos pareira L used as antidote for snake bite in indigenous medicine.* International Journal of Pharmacy and Pharmaceutical Sciences, 2014. **6**(8): p. 180-182.
124. D'Archivio, M., et al., *Polyphenols, dietary sources and bioavailability.* Annali dell'Istituto superiore di sanita, 2007. **43**(4): p. 348-61.
125. D'Archivio, M., et al., *Bioavailability of the Polyphenols: Status and Controversies.* International Journal of Molecular Sciences, 2010. **11**(4): p. 1321-1342.
126. Manach, C., et al., *Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies.* American Journal of Clinical Nutrition, 2005. **81**(1): p. 230S-242S.
127. Scalbert, A., et al., *Dietary polyphenols and the prevention of diseases.* Critical Reviews in Food Science and Nutrition, 2005. **45**(4): p. 287-306.
128. Nishihara, M. and T. Nakatsuka, *Genetic engineering of flavonoid pigments to modify flower color in floricultural plants.* Biotechnology Letters, 2011. **33**(3): p. 433-441.
129. de Groot, H. and U. Rauen, *Tissue injury by reactive oxygen species and the protective effects of flavonoids.* Fundamental & Clinical Pharmacology, 1998. **12**(3): p. 249-255.
130. Serrano, J., et al., *Tannins: Current knowledge of food sources, intake, bioavailability and biological effects.* Molecular Nutrition & Food Research, 2009. **53**(S2): p. S310-S329.
131. Wilson, J.A. and E.J. Kern, *The true tanning value of vegetable tanning materials.* Journal of Industrial and Engineering Chemistry-Us, 1920. **12**: p. 465-469.
132. Bate-Smith and Swain, *Flavonoid compounds* 1962, Comparative biochemistry III Florkin M., Mason H.S: New York. p. 75-809.
133. Haslam, E., *Plant Polyphenols, Vegetable Tannins Revisited*. 1989, Cambridge: Press Syndicate of the University of Cambridge

134. Wursch, P., et al., *The tannin granules from ripe carob pod.* Lebensmittel-Wissenschaft & Technologie, 1984. **17**(6): p. 351-354.
135. Haslam, E. and T.H. Lilley, *Natural astringency in foodstuffs-a molecular interpretation.* Critical Reviews in Food Science and Nutrition, 1988. **27**(1): p. 1-40.
136. Souquet, J.M., et al., *Polymeric proanthocyanidins from grape skins.* Phytochemistry, 1996. **43**(2): p. 509-512.
137. Sauracalixto, F., *Dietary fiber complex in a sample rich in condensed tannins and uronic-acids.* Food Chemistry, 1987. **23**(2): p. 95-103.
138. Sanoner, P., et al., *Polyphenol profiles of French cider apple varieties (Malus domestica sp.).* Journal of Agricultural and Food Chemistry, 1999. **47**(12): p. 4847-4853.
139. Cheynier, V., *Polyphenols in foods are more complex than often thought.* American Journal of Clinical Nutrition, 2005. **81**(1): p. 223-229.
140. de Pascual-Teresa, S., C. Santos-Buelga, and J.C. Rivas-Gonzalo, *Quantitative analysis of flavan-3-ols in Spanish foodstuffs and beverages.* Journal of Agricultural and Food Chemistry, 2000. **48**(11): p. 5331-5337.
141. Gu, L.W., et al., *Concentrations of proanthocyanidins in common foods and estimations of normal consumption.* Journal of Nutrition, 2004. **134**(3): p. 613-617.
142. Renard, C.M.G.C., N. Dupont, and P. Guillermin, *Concentrations and characteristics of procyanidins and other phenolics in apples during fruit growth.* Phytochemistry, 2007. **68**(8): p. 1128-1138.
143. Saura-Calixto, F., J. Serrano, and I. Goni, *Intake and bioaccessibility of total polyphenols in a whole diet.* Food Chemistry, 2007. **101**(2): p. 492-501.
144. Okuda, T., T. Yoshida, and T. Hatano, *New methods of analyzing tannins.* Journal of Natural Products, 1989. **52**(1): p. 1-31.
145. Haslam, E. and Y. Cai, *Plant polyphenols (vegetable tannins) - gallic acid metabolism.* Natural Product Reports, 1994. **11**(1): p. 41-66.
146. Hagerman, A.E. and L.G. Butler, *Determination of protein in tannin-protein precipitates.* Journal of Agricultural and Food Chemistry, 1980. **28**(5): p. 944-947.
147. Hagerman, A.E. and L.G. Butler, *The specificity of proanthocyanidin-protein interactions.* Journal of Biological Chemistry, 1981. **256**(9): p. 4494-4497.
148. Hagerman, A.E. and K.M. Klucher, *Tannin-protein interactions.* Progress in clinical and biological research, 1986. **213**: p. 67-76.

149. Baxter, N.J., et al., *Multiple interactions between polyphenols and a salivary proline-rich protein repeat result in complexation and precipitation.* Biochemistry, 1997. **36**(18): p. 5566-5577.
150. Jobstl, E., et al., *Noncovalent cross-linking of casein by epigallocatechin gallate characterized by single molecule force microscopy.* Journal of Agricultural and Food Chemistry, 2006. **54**(12): p. 4077-4081.
151. Oh, H.I., et al., *Hydrophobic interaction in tannin-protein complexes.* Journal of Agricultural and Food Chemistry, 1980. **28**(2): p. 394-398.
152. Batesmith, E.C., *Flavonoid compounds in foods.* Advances in Food Research, 1954. **5**: p. 261-300.
153. Loomis, W.D., *Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles.* Methods in enzymology, 1974. **31**: p. 528-44.
154. Haslam, E., *Vegetable Tannins*, in *Biochemistry of Plant Phenolics*, T. Swain, J. Harbone, and C. Van Sumere, Editors. 1979, Springer US. p. 475-523.
155. Kalyanaraman, B., P.I. Premovic, and R.C. Sealy, *Semiquinone anion radicals from addition of amino acids, peptides, and proteins to quinones derived from oxidation of catechols and catecholamines. An ESR spin stabilization study.* Journal of Biological Chemistry, 1987. **262**(23): p. 11080-11087.
156. Beart, J.E., T.H. Lilley, and E. Haslam, *Plant polyphenols - secondary metabolism and chemical defense - some observations.* Phytochemistry, 1985. **24**(1): p. 33-38.
157. Santos-Buelga, C. and V. Freitas, *Influence of Phenolics on Wine Organoleptic Properties*, in *Wine Chemistry and Biochemistry*, M.V. Moreno-Arribas and M.C. Polo, Editors. 2009, Springer New York. p. 529-570.
158. Oh, H.I. and J.E. Hoff, *pH Dependence of Complex Formation Between Condensed Tannins and Proteins.* Journal of Food Science, 1987. **5**(52): p. 1267–1269.
159. Soares, S., et al., *Effect of Condensed Tannins Addition on the Astringency of Red Wines.* Chemical Senses, 2012. **37**(2): p. 191-198.
160. Maury, C., et al., *Influence of fining with plant proteins on proanthocyanidin composition of red wines.* American Journal of Enology and Viticulture, 2003. **54**(2): p. 105-111.
161. Jobstl, E., et al., *Molecular model for astringency produced by polyphenol/protein interactions.* Biomacromolecules, 2004. **5**(3): p. 942-949.
162. Pascal, C., et al., *Study of the Interactions Between a Proline-Rich Protein and a Flavan-3-ol by NMR: Residual Structures in the Natively Unfolded Protein*

Provides Anchorage Points for the Ligands. Biopolymers, 2009. **91**(9): p. 745-756.

163. Wroblewski, K., et al., *The molecular interaction of human salivary histatins with polyphenolic compounds.* European Journal of Biochemistry, 2001. **268**(16): p. 4384-4397.
164. Hagerman, A.E., M.E. Rice, and N.T. Ritchard, *Mechanisms of protein precipitation for two tannins, pentagalloyl glucose and epicatechin(16) (4 -> 8) catechin (procyanidin).* Journal of Agricultural and Food Chemistry, 1998. **46**(7): p. 2590-2595.
165. Butler, L.G., et al., *Interaction of proteins with sorghum tannin - mechanism, specificity and significance.* Journal of the American Oil Chemists Society, 1984. **61**(5): p. 916-920.
166. de Freitas, V. and N. Mateus, *Structural features of procyanidin interactions with salivary proteins.* Journal of Agricultural and Food Chemistry, 2001. **49**(2): p. 940-945.
167. Soares, S., N. Mateus, and V. De Freitas, *Interaction of different polyphenols with bovine serum albumin (BSA) and human salivary alpha-amylase (HSA) by fluorescence quenching.* Journal of Agricultural and Food Chemistry, 2007. **55**(16): p. 6726-6735.
168. Poncet-Legrand, C., et al., *Flavan-3-ol Aggregation in Model Ethanolic Solutions: Incidence of Polyphenol Structure, Concentration, Ethanol Content, and Ionic Strength.* Langmuir, 2003. **19**(25): p. 10563-10572.
169. McManus, J., et al., *The Association of Proteins with Polyphenols.* Journal Chemical Society Chemical Communications, 1981: p. 309-311.
170. Charlton, A.J., E. Haslam, and M.P. Williamson, *Multiple conformations of the proline-rich protein/epigallocatechin gallate complex determined by time-averaged nuclear Overhauser effects.* Journal of the American Chemical Society, 2002. **124**(33): p. 9899-9905.
171. Siebert, K.J., *Effects of protein-polyphenol interactions on beverage haze, Stabilization and analysis.* Journal of Agricultural and Food Chemistry, 1999. **47**(2): p. 353-362.
172. Hagerman, A.E. and L.G. Butler, *The specificity of proanthocyanidin-protein interactions.* Journal of Biological Chemistry, 1981. **256**(9): p. 4494-7.
173. Asquith, T.N., et al., *Binding of condensed tannins to salivary proline-rich glycoproteins - the role of carbohydrate.* Journal of Agricultural and Food Chemistry, 1987. **35**(3): p. 331-334.

174. Abd El Mohsen, M.M., et al., *Uptake and metabolism of epicatechin and its access to the brain after oral ingestion.* Free Radical Biology and Medicine, 2002. **33**(12): p. 1693-1702.
175. Baba, S., et al., *Absorption and urinary excretion of (-)-epicatechin after administration of different levels of cocoa powder or (-)-epicatechin in rats.* Journal of Agricultural and Food Chemistry, 2001. **49**(12): p. 6050-6056.
176. Donovan, J.L., et al., *Catechin is metabolized by both the small intestine and liver of rats.* Journal of Nutrition, 2001. **131**(6): p. 1753-1757.
177. Kuhnle, G., et al., *Epicatechin and catechin are O-methylated and glucuronidated in the small intestine.* Biochemical and Biophysical Research Communications, 2000. **277**(2): p. 507-512.
178. Gonthier, M.P., et al., *Metabolism of dietary procyanidins in rats.* Free Radical Biology and Medicine, 2003. **35**(8): p. 837-844.
179. Spencer, J.P.E., et al., *Decomposition of cocoa procyanidins in the gastric milieu.* Biochemical and Biophysical Research Communications, 2000. **272**(1): p. 236-241.
180. Rios, L.Y., et al., *Cocoa procyanidins are stable during gastric transit in humans'.* American Journal of Clinical Nutrition, 2002. **76**(5): p. 1106-1110.
181. Tsang, C., et al., *The absorption, metabolism and excretion of flavan-3-ols and procyanidins following the ingestion of a grape seed extract by rats.* British Journal of Nutrition, 2005. **94**(2): p. 170-181.
182. Goel, G., et al., *Interaction of gut microflora with tannins in feeds.* Naturwissenschaften, 2005. **92**(11): p. 497-503.
183. Butler, L.G. and J.C. Rogler, *Biochemical-mechanisms of the antinutritional effects of tannins.* Acs Symposium Series, 1992. **506**: p. 298-304.
184. Kahle, K., et al., *Polyphenols are intensively metabolized in the human gastrointestinal tract after apple juice consumption.* Journal of Agricultural and Food Chemistry, 2007. **55**(26): p. 10605-10614.
185. Deprez, S., et al., *Polymeric proanthocyanidins are catabolized by human colonic microflora into low-molecular-weight phenolic acids.* Journal of Nutrition, 2000. **130**(11): p. 2733-2738.
186. Smith, A.H. and R.I. Mackie, *Effect of condensed tannins on bacterial diversity and metabolic activity in the rat gastrointestinal tract.* Applied and Environmental Microbiology, 2004. **70**(2): p. 1104-1115.
187. Daniel, E.M., et al., *The effects of ph and rat intestinal contents on the liberation of ellagic acid from purified and crude ellagitannins.* Journal of Natural Products, 1991. **54**(4): p. 946-952.

188. Clifford, M.N. and A. Scalbert, *Ellagitannins – nature, occurrence and dietary burden.* J Sci Food Agric, 2000(80): p. 1118-1125
189. Haslam, E., *Natural Polyphenols (Vegetable Tannins) as Drugs: Possible Modes of Action.* Journal of Natural Products, 1996. **59**(2): p. 205-215.
190. de Freitas, V.A.P., et al., *Characterisation of oligomeric and polymeric procyanidins from grape seeds by liquid secondary ion mass spectrometry.* Phytochemistry, 1998. **49**(5): p. 1435-1441.
191. de Freitas, V.A.P., Y. Glories, and M. Laguerre, *Incidence of molecular structure in oxidation of grape seed procyanidins.* Journal of Agricultural and Food Chemistry, 1998. **46**(2): p. 376-382.
192. Gonzalez-Manzano, S., et al., *Influence of the degree of polymerisation in the ability of catechins to act as anthocyanin copigments.* European Food Research and Technology, 2008. **227**(1): p. 83-92.
193. Delcour, J.A., D. Ferreira, and D.G. Roux, *Synthesis of condensed tannins .9. the condensation sequence of leucocyanidin with (+)-catechin and with the resultant procyanidins.* Journal of the Chemical Society-Perkin Transactions 1, 1983(8): p. 1711-1717.
194. Geissman, T.A. and Yoshimur.Nn, *Synthetic proanthocyanidin.* Tetrahedron Letters, 1966(24): p. 2669-&.
195. Papadopoulou, A., R.J. Green, and R.A. Frazier, *Interaction of flavonoids with bovine serum albumin: A fluorescence quenching study.* Journal of Agricultural and Food Chemistry, 2005. **53**(1): p. 158-163.
196. Goncalves, R., et al., *Mechanisms of Tannin-Induced Trypsin Inhibition: A Molecular Approach.* Langmuir, 2011. **27**(21): p. 13122-13129.
197. Ferreira, M. and P. Gameiro, *Ciprofloxacin metalloantibiotic: an effective antibiotic with an influx route strongly dependent on lipid interaction?* The Journal of membrane biology, 2015. **248**(1): p. 125-36.
198. Wahlund, K.G., et al., *Size characterisation of wheat proteins, particularly glutenin, by asymmetrical flow field flow fractionation.* Journal of Cereal Science, 1996. **23**(2): p. 113-119.
199. Bottomley, R.C., H.F. Kearns, and J.D. Schofield, *Characterization of wheat-flour and gluten proteins using buffers containing sodium dodecyl-sulfate.* Journal of the Science of Food and Agriculture, 1982. **33**(5): p. 481-491.
200. Stevenson, S.G., et al., *Characterization of polymeric wheat proteins by flow field-flow fractionation/MALLS.* Journal of Liquid Chromatography & Related Technologies, 2003. **26**(17): p. 2771-2781.

201. Popineau, Y. and F. Pineau, *Fractionation of wheat gliadins by ion-exchange chromatography on sp trisacryl-M.* Lebensmittel-Wissenschaft & Technologie, 1985. **18**(2): p. 133-135.
202. Bean, S.R. and G.L. Lookhart, *Faster capillary electrophoresis separation of wheat proteins through modifications to buffer composition and sample handling.* Electrophoresis, 1998. **19**(18): p. 3190-3198.
203. Slađana Zilić, M.B., Mirjana Pešić , Dejan Dodig and Dragana Ignjatović-Micić, *Characterization of proteins from grain of different bread and durum wheat genotypes.* Int J Mol Sci 12 (9), 2011: p. 5878-5894.
204. Mejías, J.H., Lu, X., Osorio, C., Jeffrey L. Ullman, von Wettstein, D., Rustgi, S., *Analysis of wheat prolamins, the causative agents of celiac sprue, using reversed phase high performance liquid chromatography (RP-HPLC) and matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS-MS).* Nutrients 6 (4), 2014: p. 1578-1597.
205. Shewry, P.R. and N.G. Halford, *Cereal seed storage proteins: Structures, properties and role in grain utilization.* J. Exp. Bot., 53, 2002: p. 947–958.
206. Triboï, E., P. Martre, and A.M. Triboï-Blondel, *Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content.* J. Exp. Bot., 54, 2003: p. 1731–1742.
207. Real, A., et al., *Identification and In Vitro Reactivity of Celiac Immunoactive Peptides in an Apparent Gluten-Free Beer.* Plos One, 2014. **9**(6).
208. Goncalves, R., N. Mateus, and V. De Freitas, *Biological Relevance of the Interaction between Procyanidins and Trypsin: A Multitechnique Approach.* Journal of Agricultural and Food Chemistry, 2010. **58**(22): p. 11924-11931.
209. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy*. 2nd ed. ed. 1999, New York: Kluwer Academic/Plenum Publishers.
210. Goncalves, R., N. Mateus, and V. de Freitas, *Inhibition of alpha-amylase activity by condensed tannins.* Food Chemistry, 2011. **125**(2): p. 665-672.
211. Rasoulzadeh, F., et al., *Fluorescence quenching study of quercetin interaction with bovine milk xanthine oxidase.* Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy, 2009. **72**(1): p. 190-193.
212. Charlton, A.J., et al., *Polyphenol/peptide binding and precipitation.* Journal of Agricultural and Food Chemistry, 2002. **50**(6): p. 1593-1601.
213. Canon, F., et al., *Aggregation of the Salivary Proline-Rich Protein IB5 in the Presence of the Tannin EgCG.* Langmuir, 2013. **29**(6): p. 1926-1937.

SUPLEMENTARY INFORMATION

S1 IDENTIFICATION OF PROTEINS CONTAINED IN THE RAW EXTRACT BY NANO-ESI-MS/MS

S1.1 BSA Tryptic Digestion Analysis

Bovine Serum Albumin (BSA) was tested as a control. *Figure S1* shows the Nano-LC-MS chromatographic profile of the resulting tryptic digestion after SDS-PAGE band excision.

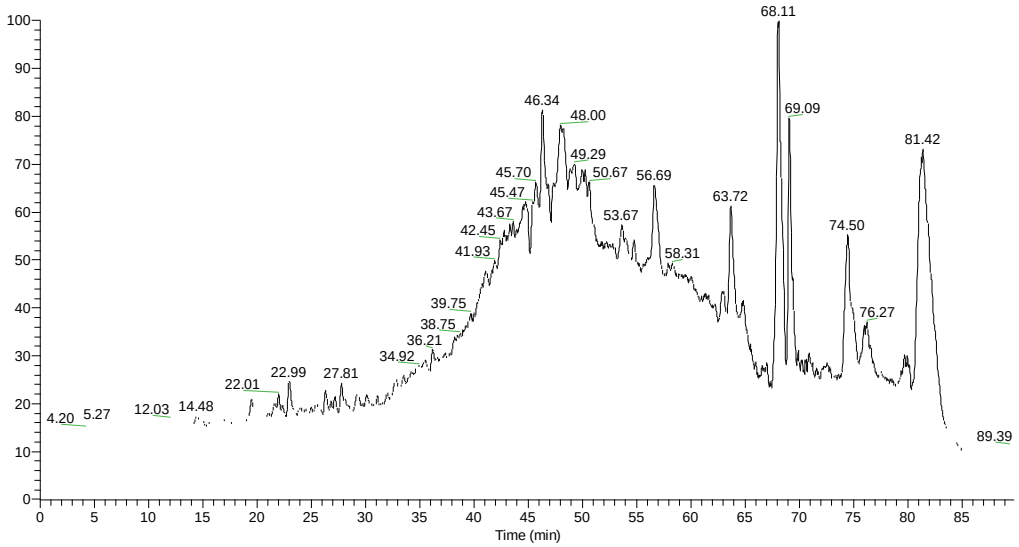
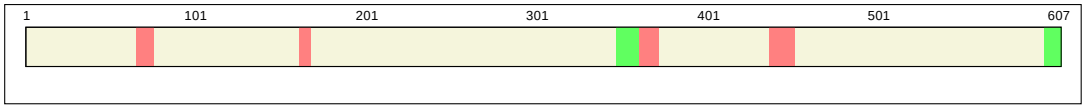


Fig. S1 - Nano-LC-MS chromatogram of BSA tryptic digest.

After analizing the MS/MS spectra, the protein identified with higher score (3.43%) was the coded as P02769 namely serum albumin. The coverage gotten was 11.04% as follows:

Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4 - [ALBU_BOVIN]



MKWVTFISLL LLFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVK**LVNEL TEFAK**TCVAD ESHAGCEKSL
HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK PDPNTLCDEF KADEKKFWGK **YLYEIAR**RHP YFYAPELLYY
ANKYNGVFQE CCQAEDKGAC LLPKIETMRE KVLASSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE FVEVTKLVTD LTKVHKECCH
GDLLECADDR ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAK**DAFL GSFLYEYSR**R
HPEYAVSVLL RLAKEYEATL EECCAKDDPH ACYSTVFDKL KHLVDEPQNL IKQNCDQFEK LGEYGFQNAL IVRYTR**KVPQ VSTPTLVEVS**
RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP
DTEKQIKKQT ALVELLKHKP KATEEQLKTV MENFVAFVDK CCAADDKEAC FAVEGPK**LVV STQTALA**

Table S1 - BSA sequenced peptides by Nano-LC-ESI-MS/MS.

PEPTIDE	RT (MIN)	MW (KDA)
LVNELTEFAK	26.44	1163.62944
YLYEIAR	22.94	927.49199
DAFLGSFLYEYSR	36.70	1567.74224
RHPEYAVSVLLR	23.29	1439.81302
KVPQVSTPTLVEVSR	23.34	1639.94438
LVVSTQTALA	25.41	1002.58232

Moreover, other three proteins were identified as a part of the sample analyzed. Glyceraldehyde-3-phosphate dehydrogenase [G3P_BOVIN] with a score of 2.76 and a coverage of 6.31% was identified through the peptide WGDAGAEYVVESTGVFTTMEK. Vimentin [VIME_BOVIN] was identified in the sample with a 2.52 score value and a coverage of 9.23% through the sequencing of LLQDSVDFSLADAINTEFK and RMFGGPGTASRPSSTRSYVTTSTR peptides. 60S ribosomal protein [RL37A_BOVIN] was also found with a 1.71 score. The protein coverage reached the 19.57% by the peptide sequenced as TVAGGAWTYNTTSAVTVK.

In summary, when the spectrum displayed in *Fig. S1* was uploaded to Proteome Discoverer 1.4 software, four different hypothesis for protein identity were given as being the most probable. These options were hypothesized with different score values and protein coverages. The molecular weight of each hypothesis was also provided. The choice of the correct protein identity was made considering the molecular weight of each SDS-PAGE excised band as well as the magnitude of obtained score values.

During sample analysis, several peaks corresponding to non-proteins were detected. At 68.11 minutes, for example, the typical PEG (polietilenglicol) mass spectrum was observed (*Fig. S2*).

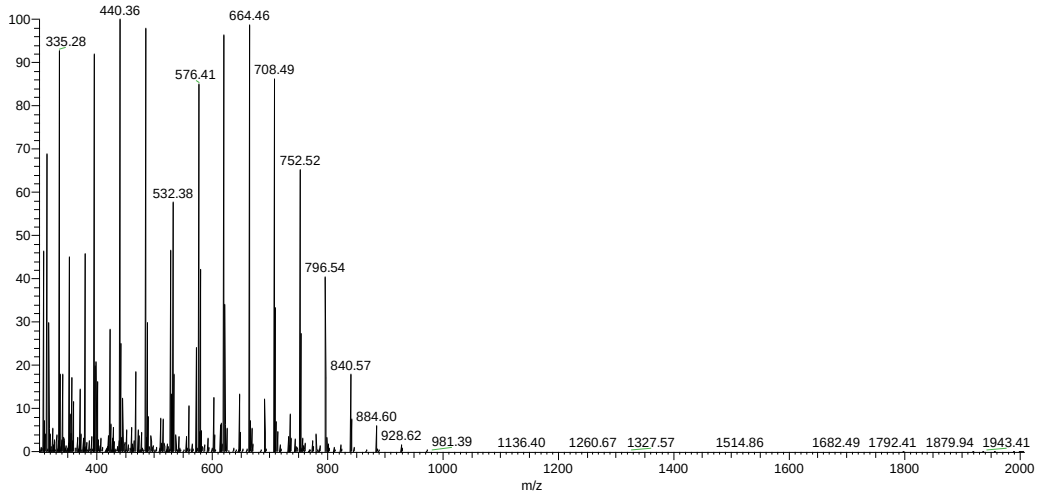


Fig. S2 - Nano-LC-MS spectra of PEG.

Besides PEG, other impurities were additionally found during column cleaning. This arised from the high list of contaminants derived from sample preparation, mainly plastic by-products and polysiloxanes, coupled to a high sensitivity of the equipment. Concerning the latter compounds, they were eluted not only during the chromatographic cleaning step but also within each run. Indeed, siloxanes are often present in the laboratory air. These compounds, which exist in deodorants and other cosmetic products usually appear in ESI-MS analysis, especialy if performed at nano conditions. To account for these common interferences, polysiloxanes were used as mass references. Given that the last peaks corresponded to ubiquous environmental contaminants, only the first 65 min of each run were considerated for interpretation. Once it has been verified the efficiency of the applied approach, the wheat gliadins raw extract was also analized by Nano-LC-ESI-MS/MS in order to identify its protein components.

S1.2 Wheat Gliadins Raw Extract Protein Identification

For the wheat gliadins raw extract, the total ion current (TIC) chromatogram obtained for each excised band (1' to 6') after their tryptic digestion as well as the different possibilities for protein identity are given below.

S1.2.1 Band 1'

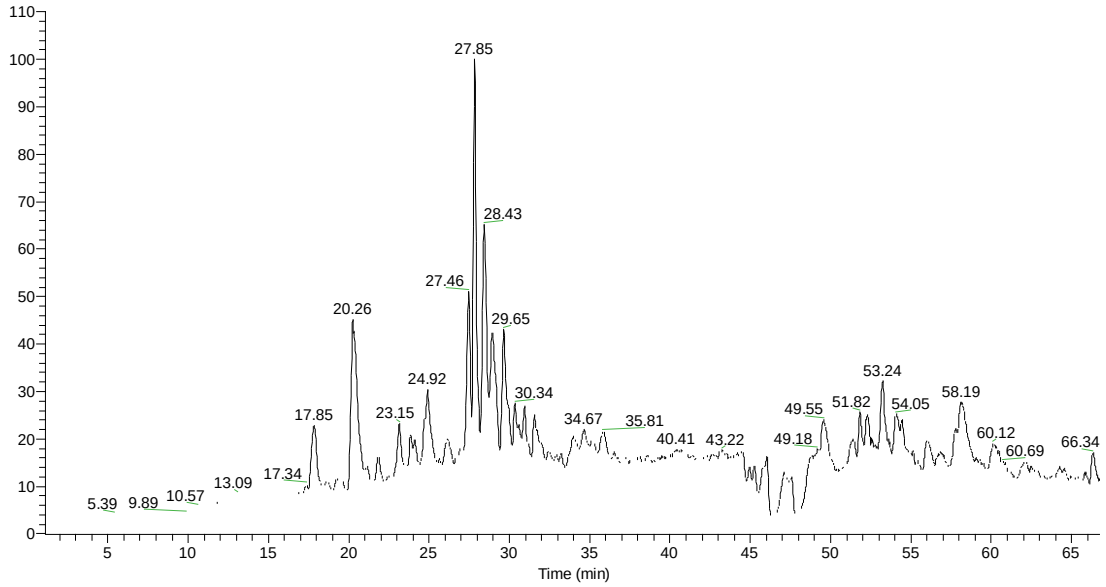
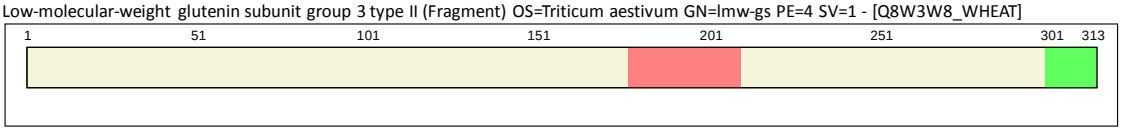


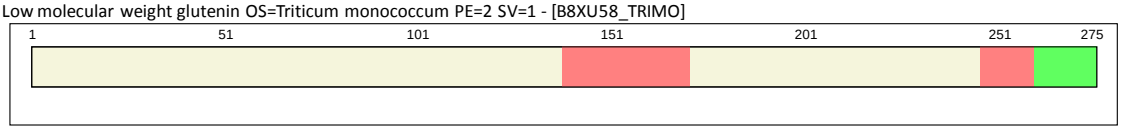
Fig. S3 - Nano-LC-MS chromatogram of band 1' tryptic digest.

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W8_WHEAT]



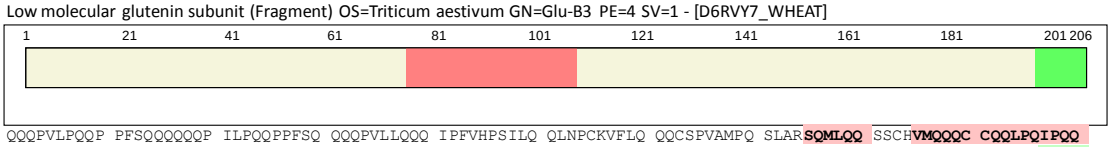
IFALLAVAAT SAIAQMENSH IPGLERPSQQ QPLPPQQTLS HHQQQQPIQQ QPHQFPQQQP CSQQQQQPPL SQQQQPPFSQ QQQPPFSQQQ QPVLPQQPPF
SQRQLPPFSQ QQQPPFSQQQ QPVLPQQPPF SQQQQPVLLQ QQIPFVHPSI LQQLNPCKVF LQQQCSPVAM PQSLAFSQML QQSSCHVMQQ QCCQQLPQIP
QQSRYEAIRA IVYSIILQEQ QQVQGSIQTQ QQQPQQLGQC VSQPQQQSQQ QLGQQPQQQQ LAQGTFLQPH QIAQLEVMTS IALRTLPTMC NVNVPLYFTT
TRVPFGVGTG VGG

Low molecular weight glutenin OS=Triticum monococcum PE=2 SV=1 - [B8XU58_TRIMO]



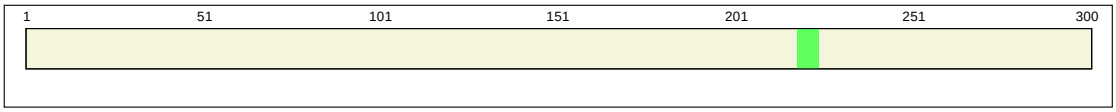
MSHIPGLEGP SQQQPLPPQQ TLTHHQQQQP IQQQPHQFPQ QQPCSQQQQQ PPLSQQQQPP FSQQQQPPFS QQQQPVLPQQ PSFSQQQQLI LPPQQQQQLP
QQQISIVQPS VLQQLNPCKV FLQQQCSPVA MPQSLARSQM LQQSSCHVMQ QQCCQQLPQI PQQSRYEAIR AIVYSIILQE QQQVQGSIQT QQQQPQQLGQ
CVSQPQQQSQ QQLGQQPQQQ QLAQGTFLQP HQIAQLEVMT SIALRTLPTM CNVNVSLYRT TTRVPFGVGT GVGGY

Low molecular glutenin subunit (Fragment) OS=Triticum aestivum GN=Glu-B3 PE=4 SV=1 - [D6RVY7_WHEAT]



QQQPVLPQQP PFSQQQQQQP ILPQQPPFSQ QQQPVLLQQQ IPFVHPSILQ QLNPCKVFLQ QQCSPVAMPQ SLARSQMLQQ SSCHVMQQQC CQQLPQIPQQ
SRYEAIRAIV YSIILQEQQQ VQGSIQTQQQ QPQQLGQCVS QPQQQSQQQL GQQPQQQQLA QGTFLQPHQI AQLEVMTSIA LRTLPTMCNV NVPLYRTTTR
VPFGVG

Alpha-gliadin OS=Triticum aestivum GN=gli-2 PE=4 SV=1 - [R9XW75_WHEAT]



MKTFLILALL AIVATTATTA VRVPVPQPQP QNPSQPQPQG QVPLVQQQQF PGQQQQFPPQ QPYPQPQPFP SQQPYLRLQP FPQPQPFPPQ
LPYPQPPPFS PQQPYPQPQPQYPQPQQPIS QQQAQQQQQQ QQQQQQQQQQ QQILQQILQQ QLIPCRDVVL QQHNIAHARS QVLQQSTYQP
LQQLCCQQLW QIPEQSRCQA IHNVVHAIIL HQQQQQRQQQ QQKQPLSQVS FQQPQQQYPS GQGSFQPSQQ NPQAQGSVQP QQLPQFEEIR
NLALETLPAM CNVYIPPYCT IAPVGIFGTN

S1.2.3 Band 2´

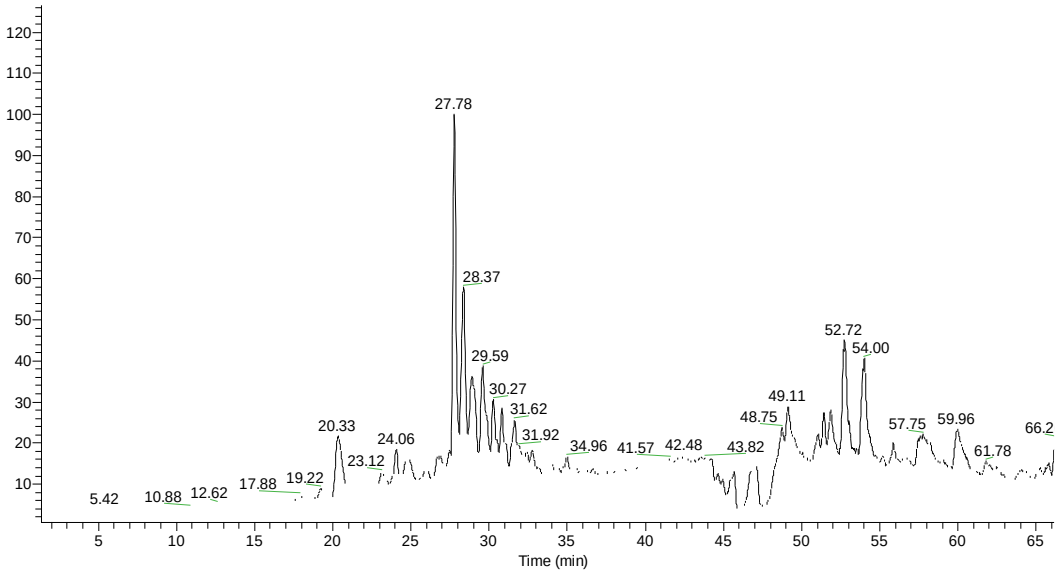
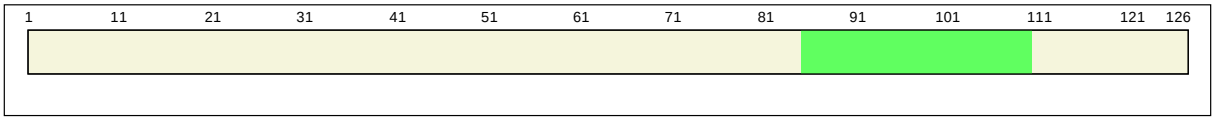


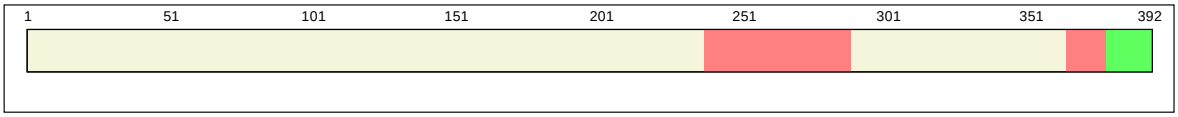
Fig. S4 - Nano-LC-MS chromatogram of band 2' tryptic digest.

Gamma-gliadin (Fragment) OS=Triticum aestivum PE=4 SV=1 - [Q1W676_WHEAT]



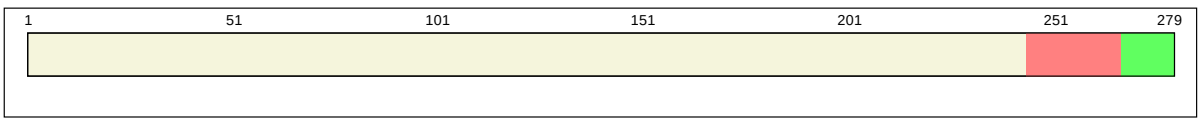
PSIQLSLQQQ MNPCKNFLLQ QCNPVSLVSS LISMILPRSD CQVMQQQCCQ QLAQIPQQLQ CAAIHSVVHS IIMQQEQRQG VQIR**RPLFQL**
VQGQGIIQPQ QPAQLEVIRS LVLRTLPTMC NVYVPR

Low molecular weight glutenin subunit OS=Triticum aestivum GN=GluB3-6 PE=4 SV=1 - [B2Y2R3_WHEAT]



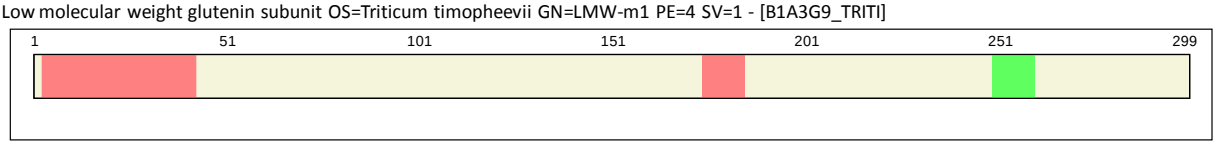
MKTFLIFALL AVAATSAIAQ MENSHIPGLE RPSQQQPLPP QQTLSHHQQQ QPIQQQPHQF PQQQPCSQQQ QQPPLSQQQQ PPFSQQQQPP FSQQQQPVLP
QQPSFSQQQL PPFSQQQQPP FSQQQQPVLP QQPSFSQQQL PPFSQQLPPF LQQQQPVLPQ QPPFSQQQLP PFSQQLPPFS QQQQPVLPQQ PPFSQQQQQP
ILPQQPPFSQ QQQPVLLQQQ IPFVHPSILQ QLNPCK**VFLQ QQCSPVAMPQ SLARSQMLQQ RSCHVMQQQC CQQLPQIPQQ SRYEAIR**AIV YSIILQEQQQ
VQGSIQTQQQ QPQQLGQCVS QPQQQLQQQL GQQPQQQQLA QGTFLQPHQI AQLEVMTSIA LR**TLPTMCNV NVPLYRTTTR VPFGVGTGVG GY**

Gamma-gliadin (Fragment) OS=Triticum aestivum PE=4 SV=1 - [B6DQB2_WHEAT]



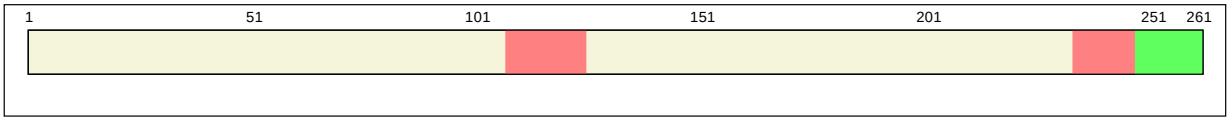
NMQVDPSGQV PWPQQQPFPQ PHQPFSQQPQ QTFPQPQQTF PHQPQQQFSQ PQQPQQQFIQ PQQPFPQQPQ QTYPQRPQQL FPQTQQPQQP FPQSQQPQQP
FPQPQQQFPQ PQQPQQSFPQ QQPSLIQQSL QQQLNPCKNL LLQQCKPVSL VSSLWSMILP RSDCQVMRQQ CCQQLAQIPQ QLQCAAIHSI VHSIIMQQEQ
QEQRQGVQIL VPLSQQQQVG QGTLVQGQGI IQPQQPAQLE VIR**SLVLQTL ATMCNVYVPP YCSTIRAPFA SIVAGIGGQ**

Low molecular weight glutenin subunit OS=Triticum timopheevii GN=LMW-m1 PE=4 SV=1 - [B1A3G9_TRITI]



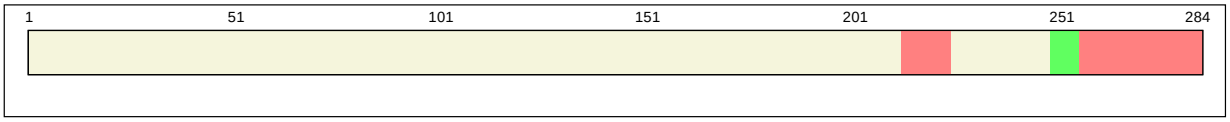
MK**TFLVFALL AVVATSAIAQ METSCIPGLG ETMGSNNHYN** KKRHFHNKPP SSQQQQPFPQ QPPFLQQQPS FSQQPLFSQK QQPVLPQQPA FSQQQQTVLP
QQPAFPQQQH QQLLQQQIPI VHPSILQQLN PCKVFLQQQC SPVAMPQHLA RSQMWQQSSC NVMQQQCCQQ LPRIPEQSRY EAIRAIIFSI ILQEQQQGFV
QPQQQQPQQS VQGVYQPQQQ SQQQLGQCSF QQPQQQLGQQ PQQQQVQKGT **FLQPHQIAR**L EVMTSIALRT LPTMCSVNVP LYSSITSAPL GVGSRVGAY

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W5_WHEAT]



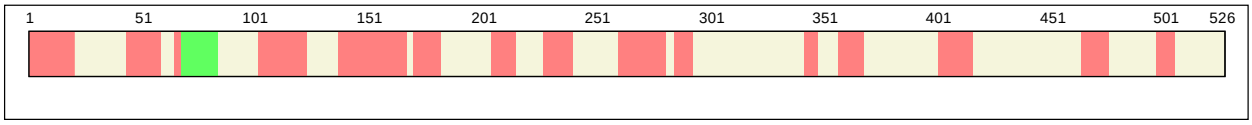
IFALLAVAAT SAIAQMENSH IPGLERPSQQ QPLPPQQTLS HHHQQQPIQQ QPHQFPQQQP CSQQQQQPPL SQQQQPPFSQ QQQPVLLQQQ IPFVHPSILQ
QLNPCK**VFLQ QQCSPVAMPQ** SLARSQMLQQ GSCHVMQQQC CQQLPQIPQQ SRYEAIRAIV YSTILQEQQQ VQGSIQTQQQ QPQQLGQCVS QPQQQSQQQL
GQQPQQQQLA QGTFLQPHQI AQLEVMTSFA LR**TLPTMCNV NVPLYRTTTR VPFGVGTGVG G**

Avenin-like b3 OS=Triticum aestivum PE=2 SV=1 - [AVLB3_WHEAT]



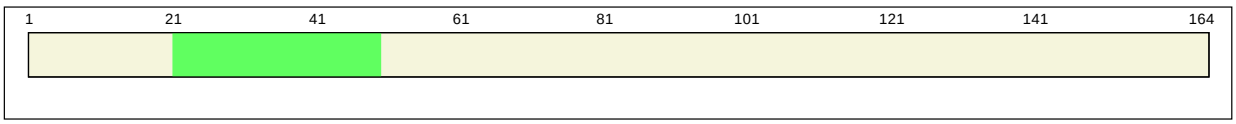
MKVFILALLA LTATTAIAQL ETTCSQGFGQ SQQQQQPGQR QLLEQMKPCV AFLQQKCSPL RMPFLQTQVE QLSSCQIVQY QCCQQLAQIP ERTRCHAIHI
VVEAIIQQQS QQQWQEPQQQ AQHKSMRMLL ENLSLMCNIY VPVQCQQQQQ LGQQQQQQLQ EQLTPCTTFL QQQCSPVTVP FPQIPVDQPT SCQNVQHQCC
RQLSQIPEQF R**CQAIHNVAE** AIRQQQPQQQ WQGMYQPQQP AQLESIR**MSL QALRSMCNIY IPVQCPAPTT YNIPLVATYT GGAC**

HMW glutenin subunit 1ByX OS=Triticum turgidum subsp. durum PE=4 SV=1 - [C3W5R6_TRITD]



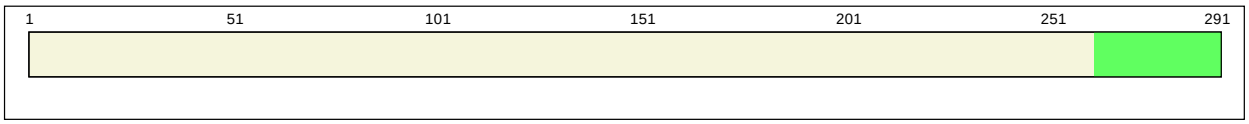
MARTGTRATR VLPNFSASVR TRTTRVLPIF SAATRTRATD RTR**ATRILPN FSAAARTR**AT DRTR**TTRVLP NFSATPRTKA TTR**TRAANRT RATTRTRAAN
R**TRATIRTRA TRVLSNFSTA AR**TRATTRTM ATIRTR**ATRV LPNFSAAART RATRAVPSFS AAARTR**ATR**A VPSFSAAART R**ATRAVPSFS AAASTRATRA
VPS**FSTTART RATR**ALPSFS AAARTR**ATTA LPSFSAATR**T RATRALHSFS AATRTRATR**A LPSFSAAGRT RTTNRTARTK** ATTR**TRATNK TR**ATTRTRAT
TRTRATSKTR ATTRTRATTR TRATRVLSNF STTVGTRATT R**TITTTR**TRA TRVLLK**FSTT ARTRATR**ALP SFSAAARTRT SRTKATTRTR ATTRTRATTR
TGATRVLSNF SAAARTRETT RTRATRVLPN FSATARTRAT TRTRATRALP NFSAADRTSA TTR**TRPTNRT SATTR**TRATR VLPNFSATVR TRATVR**TRAT
IRTR**TPTRTR AAIRTRATRL RQPIPC

HMW glutenin (Fragment) OS=Triticum aestivum GN=DAL1Sly PE=4 SV=1 - [X5CN16_WHEAT]



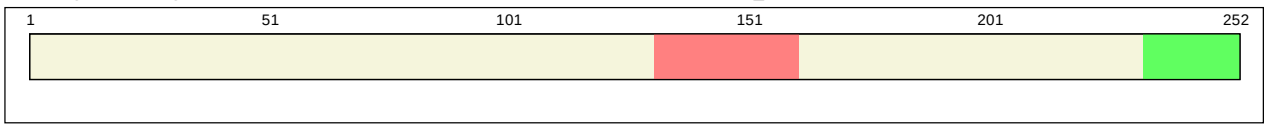
PGQGQQGYYP TSPQQPGQGK **QLGQGQQGYY PTSQQAGQGQ** QPGQGQQGRC PTSLQQTGQA QQPGQGQQIG QVQQPGQVQQ PGQGQQGYYP TSLQQPGQGQ
QSGQGQQSGQ GHQPGQEKQG YDSPYHVSAE QQAASSMVAK AQQPTTQLTT VCRMEGGDAL SASQ

Gamma-gliadin OS=Triticum aestivum PE=2 SV=1 - [B5ANT1_WHEAT]



MKTLLILTIF AAALTIATAN IQVDPSGQVQ WPQQQPFPQP QPFSQQPQQA FLQPQHTFPL QPQQVFPQPQ QPQQQFPQPQ QPQQPFPQPQ QPQLPFPQQP
QQPFPQPQQP QQPFPQSQQP QQPFPQPQQQ FPQPQQPQQS FPQQQPPLIQ PYLQQQMNPC KNYLLQQCNP VSLVSSLVSM ILPRNDCQVM QQQCCQQLAQ
IPRQLQCTAI HSVVHAIIMQ QEQQGIQILR RPLFQLVQGQ GIIQPQQPAQ YEVIRSLVLR **TLPNMFNVYV RPDCSTINAP FASIVAGIGD Q**

Gamma-gliadin (Fragment) OS=Triticum aestivum GN=GAG56D PE=4 SV=1 - [Q9FS77_WHEAT]



QVQWPQQQPV PQPHQPFSQQ PQQTFPQPQQ TFPHQPQQQF PQPQQPQQQF LQPQQPFPQQ PQQPYPQQPQ QPFPQTQQPQ QLFPQSQQPQ QQFSQPQQQF
PQPQQPQQSF PQQQPPFIQP SLQQQVNPCK **NFLLQQCKPV SLVSSLWSMI WPQSDCQVMR** QQCCQQLAQI PQQLQCAAIH TVIHSIIMQQ EQQQGMHILL
PLYQQQQVGQ GTLVQGQGII QPQQPAQLEA IR**SLVLQTLP TMCNVYVPPE CS**

S1.2.3 Band 3´

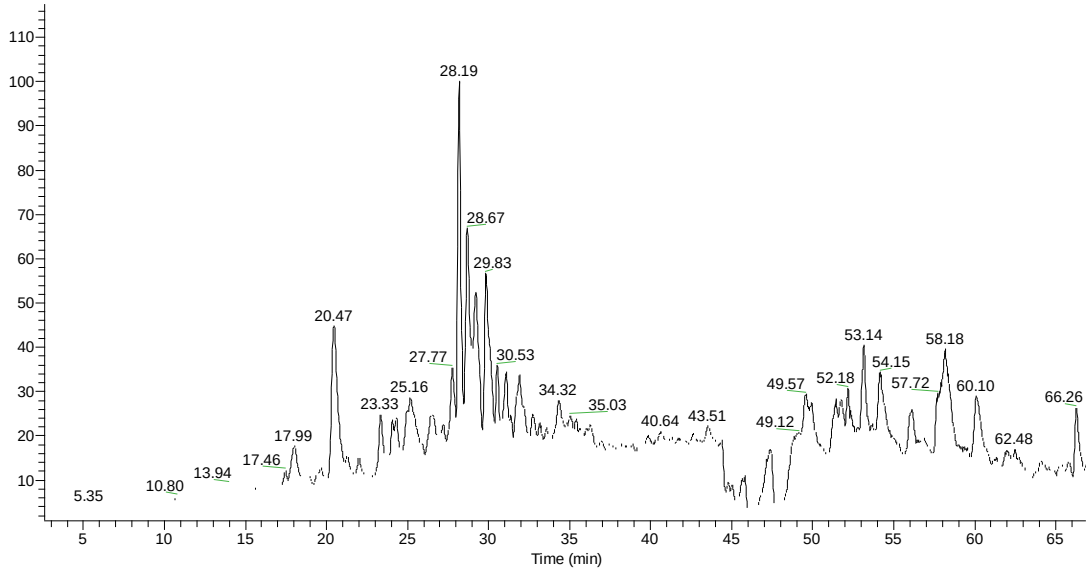
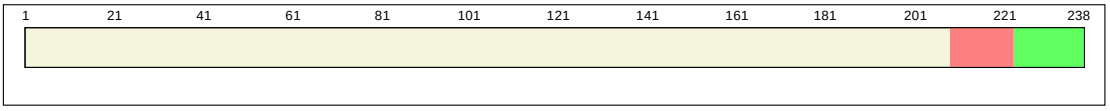


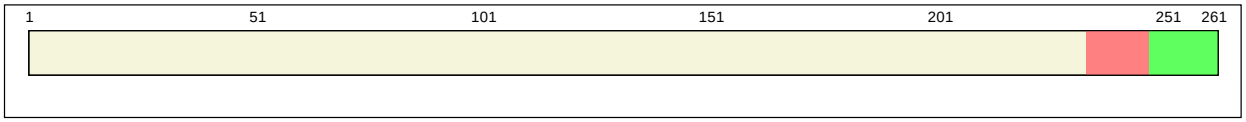
Fig. S5 - Nano-LC-MS chromatogram of band 3' tryptic digest.

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W6_WHEAT]



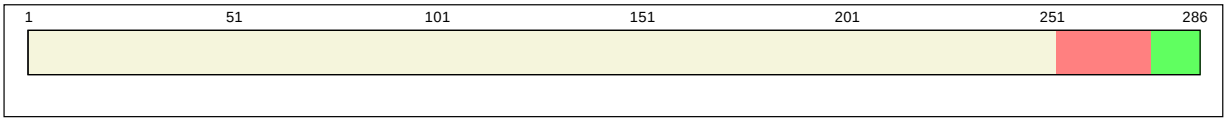
IFALLAVAAT SAIAQMENSH IPGLERLSQQ QPLPPQQTLS HHQQQQPIQQ QPHQFPQQQP CSQQQQIPFV HPSILQQLDP CKVFLQQQCS PVAMPQSLAR
SQMLQQSSCH VMQQQCCQQL PQIPQQSRYE AIRAIVYSII LQEQQQVQGS IQTQQQQPQQ LGQCVSQPQQ QSQQQLGQQP QQQQLAQGTF LQPHQIAQLE
VMTSIALR**TL PTMCNVNVPL YRTTTRVPFG VGTGVGGY**

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W5_WHEAT]



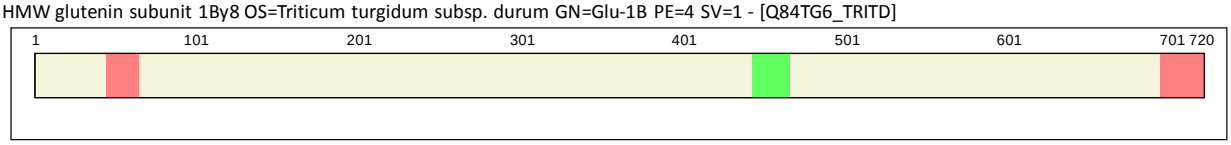
IFALLAVAAT SAIAQMENSH IPGLERPSQQ QPLPPQQTLS HHHQQQPIQQ QPHQFPQQQP CSQQQQQPPL SQQQQPPFSQ QQQPVLLQQQ IPFVHPSILQ
QLNPCKVFLQ QQCSPVAMPQ SLARSQMLQQ GSCHVMQQQC CQQLPQIPQQ SRYEAIRAIV YSTILQEQQQ VQGSIQTQQQ QPQQLGQCVS QPQQQSQQQL
GQQPQQQQLA QGTFLQPHQI AQLEVMTSFA LR**TLPTMCNV NVPLYRTTTR VPFGVGTGVG G**

Gamma gliadin OS=Triticum monococcum PE=2 SV=1 - [B8XU42_TRIMO]



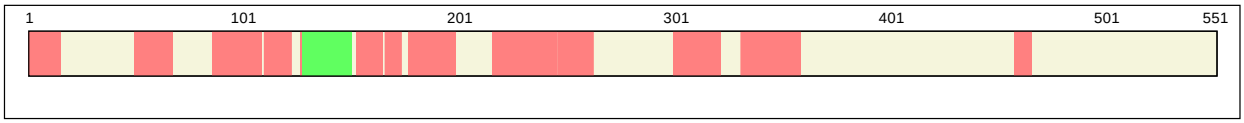
MKTLLILTIF AAALTIATAN IQVDPSGQVQ WPQQQPFPQP QPFSQQPQQA FPQPQQTFPL QPHQVFPQPQ QPQQQFPQPQ QPQQPFPQPQ QPQLPFPQQP
FPQPQQPQQP FPQSQQPQLP FLQPQQQFPQ PQQPQQSSPQ QQQPLIHPYL QQQMNPCKNY LLQQCNPVSL VSSLVSMILP RSDCQVMQQL CCQQLAQIPR
QLQCTAIHSV VHAIIMQQEQ QGIPILRLLF QLVQGQEIIQ PPTPDKNEVI R**SLELTTLPD CAKCMSDLTA PHQRFICKLV AGVAGQ**

HMW glutenin subunit 1By8 OS=Triticum turgidum subsp. durum GN=Glu-1B PE=4 SV=1 - [Q84TG6_TRITD]



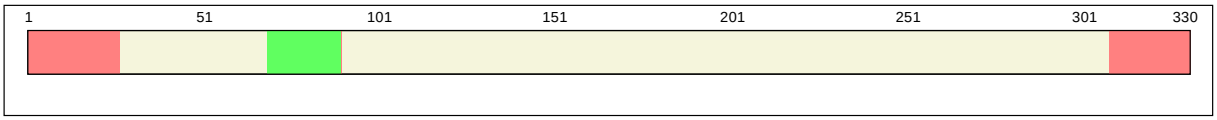
MAKRLVLFAT VVITLVALTA AEGEASRQLQ CERELQESSL EACR**QVVDQQ LAGRLPWSTG LQMR**CCQQLR DVSAKCRLVA VSQVVRQYEQ TVVPPKGGSF
YPGETTPLQQ LQQVIFWGTS SQTVQGYYPS VSSPQQGPYY PGQASPQQPG QGQQPGKWQE LGQGQQGYYP TSLHQSGQGQ QGYYPSSLQQ PGQGQQIGQG
QQGYYPTSLQ QPGQGQQIGQ GQQGYYPTSP QHPGQRQQPG QGQQIGQGQQ LGQGRQIGQG QQSGQGQQGY YPTSPQQLGQ GQQPGQWQQS GQGQQGYYPT
SQQQPGQGQQ GQYPASQQQP GQGQQGQYPA SQQQPGQGQQ GQYPASQQQP AQGQQGQYPA SQQQPGQGQQ GHYLASQQQP GQGQQRHYPA SLQQPGQGQQ
GHYTASLQQP GQGQQGHYPA SLQQVGQGQQ IGQLGQRQQP GR**GQQTRQGQ QLEQGQQPGQ GQQTR**QGQQL EQGQQPGQGQ QGYYPTSPQQ SGQGQQPGQS
QQPGQGQQGY YSSSLQQPGQ GLQGHYPASL QQPGQGHPGQ RQQPGQGQQP EQGQQPGQGQ QGYYPTSPQQ PGQGKQLGQG QQGYYPTSPQ QPGQGQQPGQ
GQQGHCPTSP QQTGQAQQPG QGQQIGQVQQ PGQGQQGYYP ISLQQSGQGQ QSGQGQQSGQ GHQLGQGQQS GQEQQGYDNP YHVNTEQQTA SPK**VAKVQQP**
ATQLPIMCRM EGGDALSASQ

Protein kinase OS=Triticum aestivum PE=3 SV=1 - [Q0Q5D1_WHEAT]



MGCCCSWIRG IRTVKCSSIR LGCCCSWVHG LRTVKCSSIR LGCCCSWVR**G SCACCCSWIS GLCGRKK**REA GKETSTSETK KTKRK**WRRSF CGWTSHEAKE PLTSETKK**KR **EAGQEASTSE TK**KTKR**KWGR GFCGWTSHEA EEPSISETKK** KR**KNGASSSE PDKK**R**WFKNK IWR**NKK**AKNK QLATLVKEIS LPNSPKAR**AA AGEILRIGNH NIPSR**VFTHS QLSDATNSFS QENLLGEGGF GRVYKGYIPE TMEVIAVKQL DK**DGLQGNRE FLVEVLMLSL LHHPNLVTLL GYCTECDQK**I LVYEYMPLGS LQDHLLDLTP K**SQPLSWHTR **MKIAVDAARG LEYLHEVANP PVVYRDLK**AS NILLDGNFSA KLADFGLAKL GPVGDKTHVT TRVMGTYGYC APEYAMSGKL TKMSDIYCFG VVFLELITGR RAIDTTKPTR EQILVHWAAP LFKDKKK**FTK MADPK**LDSKY PLKGLYQALA ISSMCLQEEA SSRPLISDVV TALTFLADPN YDPPDDIEPL PITVPDFDRG ISLGEPEEVK EKFEEKQVED S

Dof DNA-binding protein OS=Triticum aestivum PE=2 SV=1 - [Q6RK62_WHEAT]



MEEVFPSNSK SKAGQMAGEA IAGAEKKPRP KPEQKVECPR CKSGNTKFCY YNNYSMSQPR YFCKACRR**YW THGGSLRNVP IGGGCRKPK**R SGTSDAHKLG VASSPEPTTV VPPSTCTGMN FANVLPTFMS VGFEIPSSLS LTAFGSSSSS NTAAMMSPGG TTSFLDVLRG GARGLLDGSL SQNNGYYYGG PAIGSGNGML MTPPAVSFGI PVPMQQHGDL VVGGNGIGAA TASIFQGATS EEGDDGMGGV MGLQWQPQVG NGGGGGGVSG GVHHLGTGNN VTMGNSNIHN NNNNDSGGDD NNGGSSR**DCY WINNGGSNPW QSLLNSSSLM**

S1.2.4 Band 4´

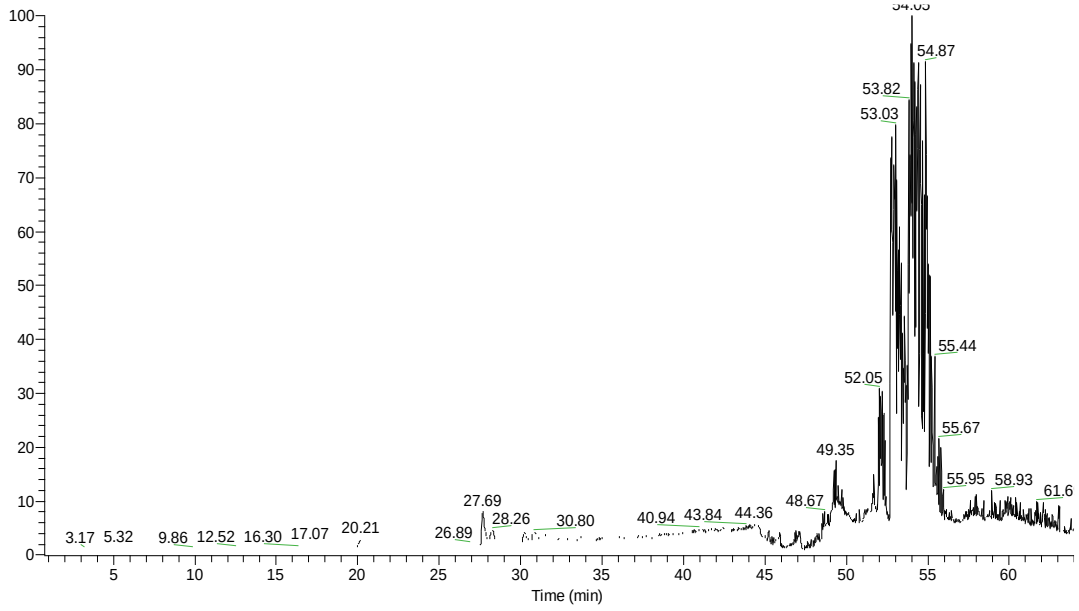
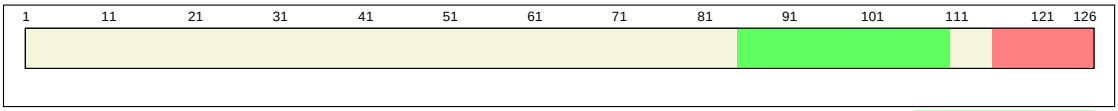


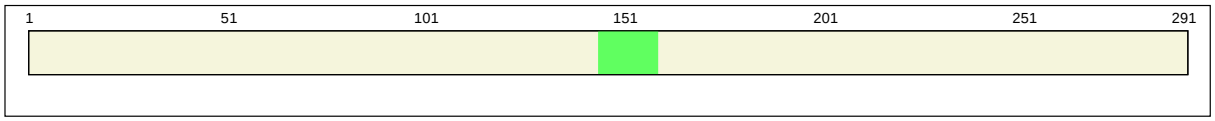
Fig. S6 - Nano-LC-MS chromatogram of band 4' tryptic digest.

Gamma-gliadin (Fragment) OS=Triticum aestivum PE=4 SV=1 - [Q1W676_WHEAT]



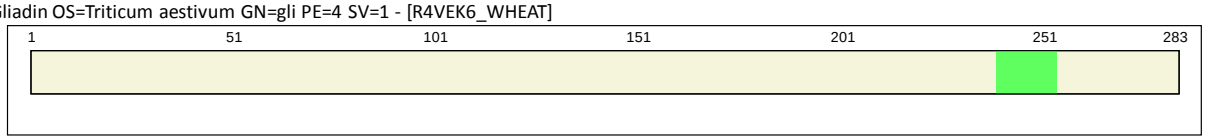
PSIQLSLQQQ MNPCKNFLLQ QCNPVSLVSS LISMILPRSD CQVMQQQCCQ QLAQIPQQLQ CAAIHSVVHS IIMQQEQRQG VQIR**RPLFQL VQGQGIIQPQ QPAQLEVIR**S LVLR**TLPTMC NVYVPR**

Alpha/beta-gliadin OS=Triticum aestivum PE=4 SV=1 - [I0IT55_WHEAT]



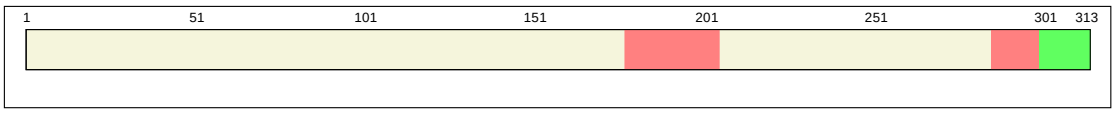
MKTFLILALL AIVATTTTTA VRVPVPQLQP QNPSQQQPQE QVPLVQQQQF LGQQQQQFPG QQQPFPPQQP YPQPQPFLPQ LPYPQPQPFP PQQSYPQPQP
QYPQPQQPIS QQQAQLQQQQ QQQQQQQQQI LQQILQQQLI PCRDVVLQQP NIAHASSKVS QQSYQLLQQL CCQQLWQTPE QSRCQAIHNV IHAIILHQQQ
QQQQQQQQQQ QQQPSSQVSY QQPQQQYPSG QGFFQPSQQN PQAQGFVQPQ QLPQFEEIRN LALQTLPAMC NVYIPPYCST TIAPFGIMST N

Gliadin OS=Triticum aestivum GN=gli PE=4 SV=1 - [R4VEK6_WHEAT]



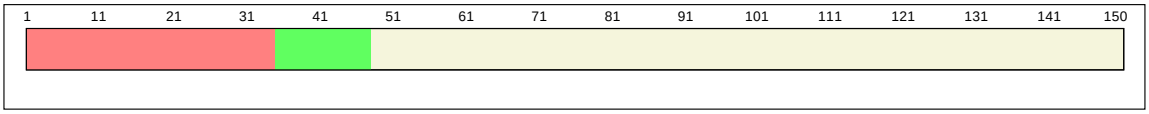
MKTFLILALL AIVATTATTA VRVPVPQLQP QHPSQQQPQE QVPLVQQQQF LGQQQPFPPQ QPYPQPQPFP SQQPYLQLQP FPQPQLPYSQ PQPFRPQQPY
PQPQPQYSQP QQPISQRQQQ QQQQQQQQQQ ILQQILQQQL IPCMDVVLQQ HNIAHGRSQV LQQSTYQLLQ ELCCQHLWQI PEQSQCQAIH NVVHAIILHQ
QQKPQQQPSS QVSFQQPLQQ YPLGQGSFRP SQQNPQARGS VQPQQLPQFE EIRNLALQTL PAMCNVYIPP YCTIAPFGIF GTN

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W8_WHEAT]



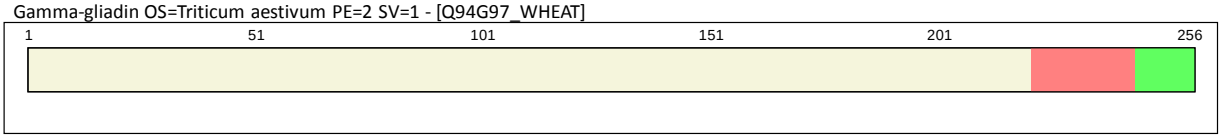
IFALLAVAAT SAIAQMENSH IPGLERPSQQ QPLPPQQTLS HHQQQQPIQQ QPHQFPQQQP CSQQQQQPPL SQQQQPPFSQ QQQPPFSQQQ QPVLPQQPPF
SQRQLPPFSQ QQQPPFSQQQ QPVLPQQPPF SQQQQPVLLQ QQIPFVHPSI LQQLNPCKVF LQQQCSPVAM PQSLARSQML QQSSCHVMQQ QCCQQLPQIP
QQSRYEAIRA IVYSIILQEQ QQVQGSIQTQ QQQPQQLGQC VSQPQQQSQQ QLGQQPQQQQ LAQGTFLQPH QIAQLEVMTS IALPTLPTMC NVNVPLYRTT
TRVPFGVGTG VGG

Alpha-gliadin (Fragment) OS=Triticum compactum PE=4 SV=1 - [F6M8E2_9POAL]



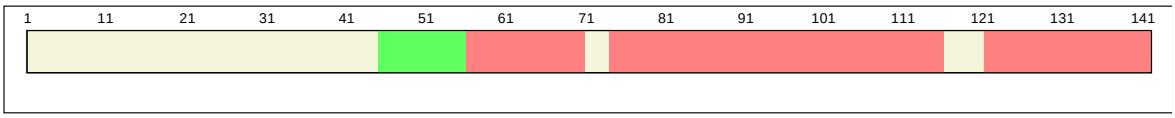
MKTFLILALL AIVATTATIA ILQQILQQQL IPCRDVVLQQ HNIAHGRSQV LQQSTYQLVQ QLCCQQLWQI PEQSRCQAIH NVVHAIILHP LTQVSFQQPQ
QQYPSGQGSF QPSQQNPQAQ GSVQPQQLPQ FEEIRNLALE TLPAMCNVYI

Gamma-gliadin OS=Triticum aestivum PE=2 SV=1 - [Q94G97_WHEAT]



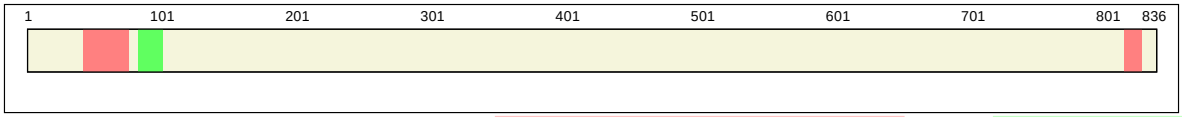
MKTLLILTII AVALTTTTAN IQVDPSGQVQ WPQQQQPFPQ PQQPQQPFPQ PQQPQLPFPQ QPQQPFPQPQ QPQQPFPQLQ QPQQPLPQPQ QPQQPFPQQQ
QPLIQPYLQQ QMNPCKNYLL QQCNPVSLVS SLVSMILPRS DCKVMRQQCC QQLARIPQQL QCAAIHGIVH SIIMQQEQQQ QQQQQQQQQQ GIQIMRPLFQ
LVQGQGIIQP QQPAQLEVIR SLVLGTLPTM CNVFVPPECS TTKAPFASIV ADIGGQ

Puroindoline a (Fragment) OS=Triticum urartu PE=4 SV=1 - [B0FHG5_TRIUA]



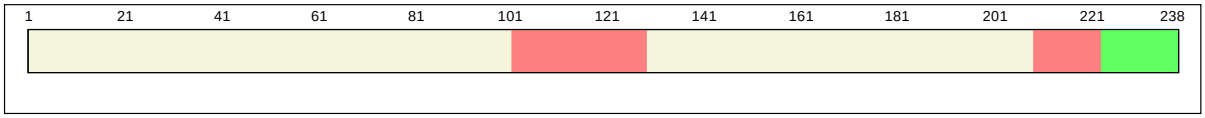
MKALFLIGLL ALVASTAFAQ YSEIVGSYDV AGGGGAQQCP LETKLNSCRN YLLGRCSTMK DFPVTWRWWK WWKGGCLELL GECCSQLGQM PPQCRCNIIQ
GSIQGDLSGI FGFQRDRASK VIQEAKTLPP RCNQGPPCDI P

High-molecular-weight glutenin subunit 1Dx2.1 OS=Triticum aestivum PE=4 SV=2 - [Q6R2V1_WHEAT]



MAKRLVLFVA VVVALVALTV AEGEASEQLQ CERELQELQE R**ELKACQQVM DQQLRDISPE CHPVVVSPVA GQYER**QIVVP PK**GGSFYPGE TTPPQQLQQR**
IFWGIPALLK RYYPSVTSPQ QVSYYPGQAS PQRPGQGQQP GQGQQSGQGQ QGYYPTSPQQ PGQWQQPEQG QPGYYPTSPQ QPGQLQQPAQ GQQPGQGQQG
QQPGQGQPGY YPTSSQLQPG QLQQPAQGQQ GQQPGQGQQG QQPGQGQQPG QGQQGQQPGQ GQQPGQGQQG QQLGQGQQGY YPTSLQQSGQ GQPGYYPTSL
QQLGQGQSGY YPTSPQQPGQ GQQPGQLQQP AQGQQPGQGQ QGQQPGQGQQ GQQPGQGQQP GQGQPGYYPT SPQQSGQGQP GYYPTSSQQP TQSQQPGQGQ
QGQQVGQGQQ AQQPGQGQQP GQGQPGYYPT SPQQSGQGQP GYYLTSPQQS GQGQQPGQLQ QSAQGQKGQQ PGQGQQPGRG QQGQQPGQGQ QGQQPGQGQP
GYYPTSPQQS GQGQQPGQWQ QPGQGQPGYY PTSPLQPGQG QPGYDPTSPQ QPGQGQQPGQ LQQPAQGQQG QQLAQGQQGQ QPAQVQQEQQ PAQGQQGQQL
GQGQQGQQPG QGQQGQQPAQ GQQGQQPGQG QQGQQPGQGQ QPGQGQPWYY PTSPQESGQG QQPGQWQQPG QGQPGYYLTS PLQLGQGQQG YYPTSLQQPG
QGQQPGQWQQ SGQGQHEYYP TSPQLSGQGQ RPGQWLQPGQ GQQGYYPTSP QQSGQGQQLG QWLQPGQGQQ GYYPTSLQQT GQGQQSGQGQ QGYYSSYHVS
VEHQAASLKV AK**AQQLAAQL PAMCR**LEGGD ALSASQ

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3W6_WHEAT]



IFALLAVAAT SAIAQMENSH IPGLERLSQQ QPLPPQQTLS HHQQQQPIQQ QPHQFPQQQP CSQQQQIPFV HPSILQQLDP CKVFLQQQCS PVAMPQSLAR
SQMLQQSSCH VMQQQCCQQL PQIPQQSRYE AIRAIVYSII LQEQQQVQGS IQTQQQQPQQ LGQCVSQPQQ QSQQQLGQQP QQQQLAQGTF LQPHQIAQLE
VMTSIALR**TL PTMCNVNVPL YRTTTRVPFG VGTGVGGY**

S1.2.5 Band 5´

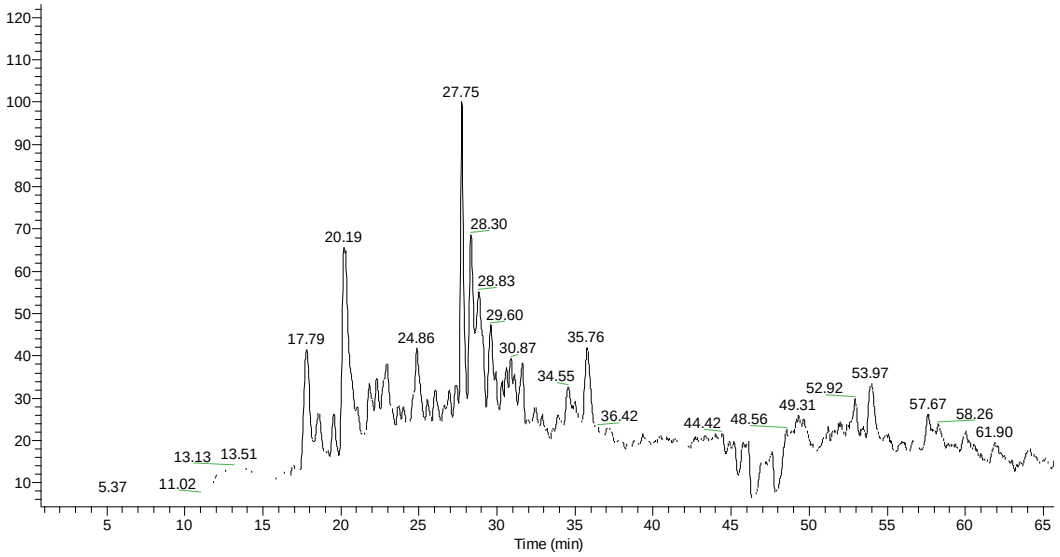
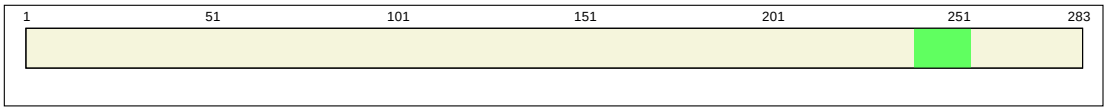


Fig. S7 - Nano-LC-MS chromatogram of band 5' tryptic digest.

Gliadin OS=Triticum aestivum GN=gli PE=4 SV=1 - [R4VEK6_WHEAT]



MKTFLILALL AIVATTATTA VRVPVPQLQP QHPSQQQPQE QVPLVQQQQF LGQQQPFPPQ QPYPQPQPFP SQQPYLQLQP FPQPQLPYSQ PQPFRPQQPY
PQPQPQYSQP QQPISQRQQQ QQQQQQQQQQ ILQQILQQQL IPCMDVVLQQ HNIAHGRSQV LQQSTYQLLQ ELCCQHLWQI PEQSQCQAIH NVVHAIILHQ
QQKPQQQPSS QVSFQQPLQQ YPLGQGSFRP SQQNPQAR**GS VQPQQLPQFE EIR**NLALQTL PAMCNVYIPP YCTIAPFGIF GTN

S1.2.6 Band 6´

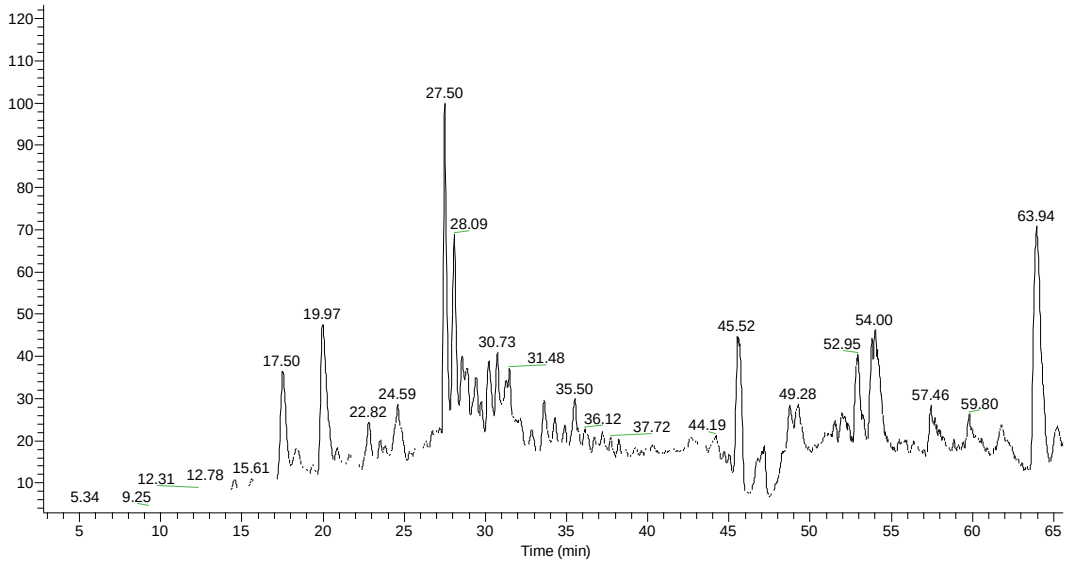
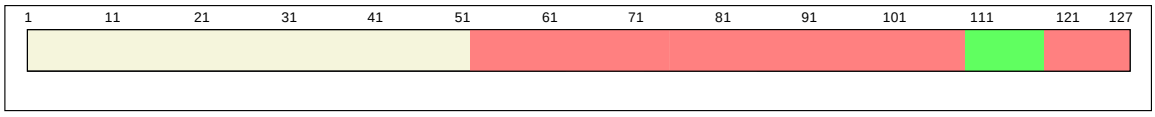


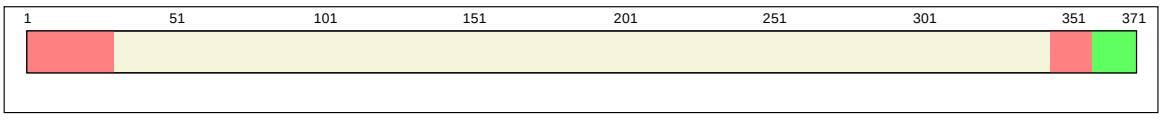
Fig. S8 - Nano-LC-MS chromatogram of band 6' tryptic digest.

Puroindolin b OS=Triticum aestivum GN=Pinb-D1ab PE=4 SV=1 - [A8R0D1_WHEAT]



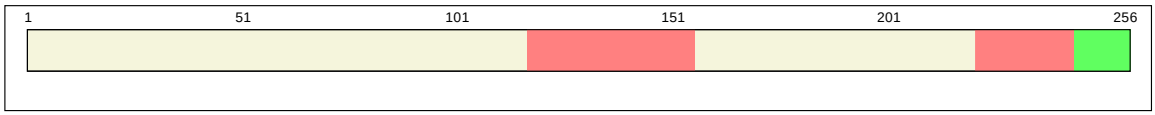
MKTLFLLALL ALVASTTFAQ YSEVGGWYNE VGGGGGSQQC PQERPKLSSC K**DYVMERCFT MKDFPVTWPT KWWKGGCEHE VREKCCKQLS QIAPQCRCDS IRRVIQGRLG GFLGIWRGEV FKQLQRA**

Low-molecular-weight glutenin subunit group 3 type II (Fragment) OS=Triticum aestivum GN=lmw-gs PE=4 SV=1 - [Q8W3X2_WHEAT]



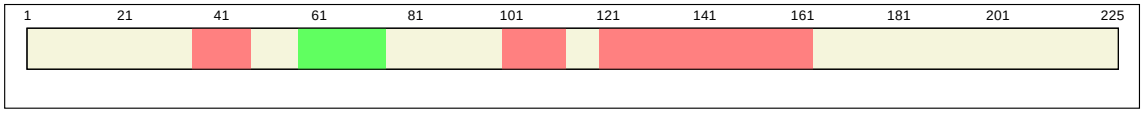
IFALLAVAAT SAIAQMENSH IPGLERPSRQ QPLPPQQTLS HHQQQQPIQQ QPHQFPQQQP CSQQQQQPPL SQQQQPPFSQ QQQPPFSQQQ QPVLPQQPSF SQQQLPPFSQ QQQPPFSQQQ QPVLPQQPSF SQQQLPPFSQ QLPPFSQQQQ PVLPQQPPFS QQQLPPFSQQ LPPFSQQQQQ VLPQQPPFSQ QQQPVLLQQQ IPFVHPSILQ QLNPCKVFLQ QQCSPVAMPQ SLARSQMLQQ SSCHVMQQQC CQQLPQIPQQ SRYEAIRAIV YSIILQEQQQ VQGSIQTQQQ QPQQLGQCVS QPQQQSQQQL GQQPQQQQLA QGTFLQPHQI AQLEVMTSIA LR**TLPTMCNV NVPLYRTTTR VPFGVGTGVG G**

Gamma-gliadin OS=Triticum aestivum PE=2 SV=1 - [Q94G97_WHEAT]



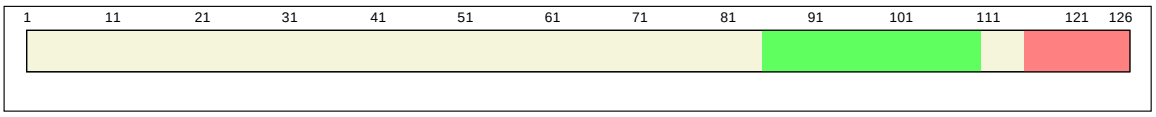
MKTLLILTII AVALTTTTAN IQVDPSGQVQ WPQQQQPFPQ PQQPQQPFPQ PQQPQLPFPQ QPQQPFPQPQ QPQQPFPQLQ QPQQPLPQPQ QPQQPFPQQQ QPLIQPYLQQ QMNPCK**NYLL QQCNPVSLVS SLVSMILPRS DCKVMRQQCC QQLAR**IPQQL QCAAIHGIVH SIIMQQEQQQ QQQQQQQQQQ GIQIMRPLFQ LVQGQGIIQP QQPAQLEVIR **SLVLGTLPTM CNVFVPPECS TTKAPFASIV ADIGGQ**

Globulin 1 OS=Triticum aestivum PE=4 SV=1 - [Q0Q5D9_WHEAT]



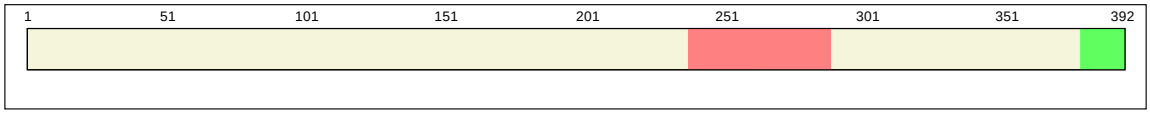
MGRFVFFALF LAALVAVSAA QGVLEQRLAD AQCR**GEVREK PLHACR**QILE QQLTGR**AGEG AFGVPLFQAQ SDAR**ERCCQQ LESVSRECRC AALRGMVR**DY EQSMPPLGEG** RHGSSGER**QP ERGCSGESTA EQRQEVQGGQ YGSETGGSQQ QGGGYHGVTV GR**GGQRQGQV LCHKRPQRQQ GEGFSGEGAQ QKPQAGRVRL TKVRLPTACR IEPQECSVFF ADQYY

Gamma-gliadin (Fragment) OS=Triticum aestivum PE=4 SV=1 - [Q1W676_WHEAT]



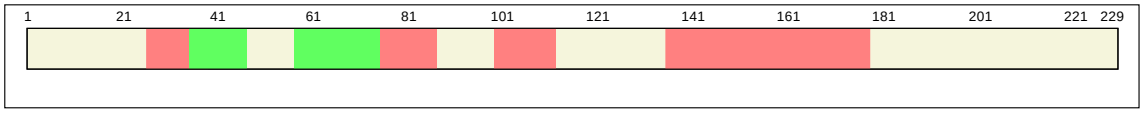
PSIQLSLQQQ MNPCKNFLLQ QCNPVSLVSS LISMILPRSD CQVMQQQCCQ QLAQIPQQLQ CAAIHSVVHS IIMQQEQRQG VQIR**RPLFQL VQGQGIIQPQ QPAQLEVIR**S LVLR**TLPTMC NVYVPR**

Low molecular weight glutenin subunit OS=Triticum aestivum GN=GluB3-6 PE=4 SV=1 - [B2Y2R3_WHEAT]



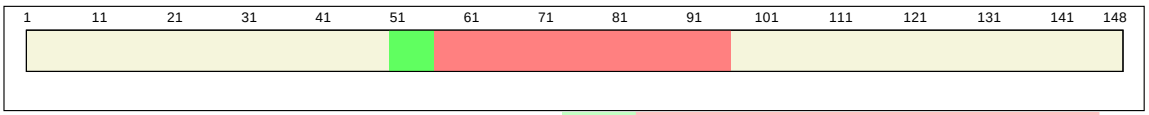
MKTFLIFALL AVAATSAIAQ MENSHIPGLE RPSQQQPLPP QQTLSHHQQQ QPIQQQPHQF PQQQPCSQQQ QQPPLSQQQQ PPFSQQQQPP FSQQQQPVLP
QQPSFSQQQL PPFSQQQQPP FSQQQQPVLP QQPSFSQQQL PPFSQQLPPF LQQQQPVLPQ QPPFSQQQLP PFSQQLPPFS QQQQPVLPQQ PPFSQQQQQP
ILPQQPPFSQ QQQPVLLQQQ IPFVHPSILQ QLNPCKVFLQ QQCSPVAMPQ SLARSQMLQQ RSCHVMQQQC CQQLPQIPQQ SRYEAIRAIV YSIILQEQQQ
VQGSIQTQQQ QPQQLGQCVS QPQQQLQQQL GQQPQQQQLA QGTFLQPHQI AQLEVMTSIA LRTLPTMCNV NVPLYPTTTR VPFGVGTGVG GY

Globulin 1 OS=Triticum aestivum PE=4 SV=1 - [Q0Q5E3_WHEAT]



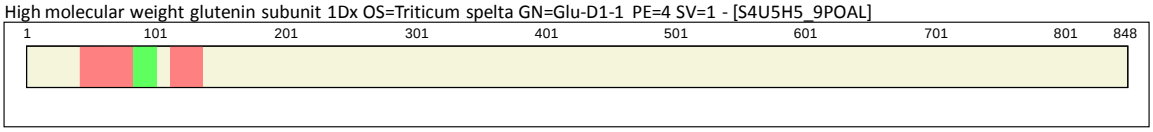
MGKFVFFAVF LAALVSVSAA QGVLKQSLTD AQCRGEVREK PLLACRQILE QQLTGRAGEG AVGVPLFQAQ WGARERCCRQ LESVSRECRC AALRGMVRDY
EQSMPPLGEG RHGSSGERQQ EQGCSGESTE PEQRQEVQGG QYGSETGGSQ QQQQGGGYHG VTVGRGGQRQ GQVLCHKRPQ RQQGEGFSGE GAQQKPQAGR
VRLTKVRLPT ACRIEPQECS VFSADQYYY

Puroindoline a OS=Triticum aestivum GN=Pina-D1 PE=4 SV=1 - [A6N860_WHEAT]



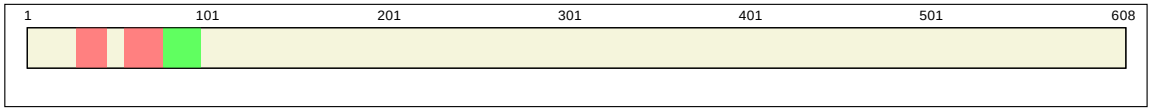
MKALFLIGLL ALVASTAFAQ YSEVVGSYDV AGGGGAQQCP VETKLNSCRN YLLDRCSTMK DFSVTWRWWK WWKGGCQELL GECCSRLGQM PPQCRCNIIQ
GSIQGDLGGI FGFQRDRASK VIQEAKNLPP RCNQGPPCNI PGTIGYYW

High molecular weight glutenin subunit 1Dx OS=Triticum spelta GN=Glu-D1-1 PE=4 SV=1 - [S4U5H5_9POAL]



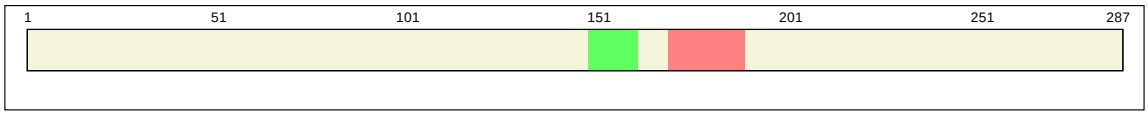
MAKRLVLFVA VVVALVALTV AEGEASEQLQ CERELQELQE RELKACQQVM DQQLRDISPE CHPVVVSPVA GQYEQQIVVP PKGGSFYPGE TTPPQQLQQR
IFWGIPALLK RYYPSVTCPQ QVSYYPGQAS PQRPGRGQQP GQGQQGYYPT SPQQPGQWQQ PEQGQPRYYP TSPQQSGQLQ QPAQGQQPGQ GQQGQQPGQG
QPGYHPTSSQ LQLGQLQQPA QGQQGQQPGQ GQQGQQPGQG QQPGQGQQGQ QPGQGQQPGQ GQQGQQLRQG QQGYYPTSLQ QSGQGQPGYY PTSLQQLGQG
QSGYYPTSPQ QPGQGQQPGQ LQQPAQGQQP GQGQQGQQPG QGQQGQQPGQ GQQPGQGQPG CYPTSPQQSG QGQPGYYPTS SQQPTQSQQP GQGQQGQQVG
QGQQAQQPGQ GQQPGQGQPG YYPTSPQQSG QGQPGYYLTS PQQSGQGQQP GQLQQSAQGQ KGQQPGQGQQ PGQGQQGQQP GQGQQGQQPG QGQPGYYPTS
PQQSGQGQQP GQWQQPGQRQ PGYYPTSPLQ PGQGQPGYDP TSPQQPGQGQ QPGQLQQPAQ GQQGQQLAQG QQGQQPAQVQ QGQRPAQGQQ GQQPGQGQQG
QQLGQGQQGQ QPGQGQQGQQ PAQGQQGQQP GQGQQGQQPG QGQQGQQPGQ GQQPGQGQPW YYPTSPQESG QGQQPGQWQQ PGQGQPGYYL TSPLQLGQGQ
QGYYPTSLQQ PGQGQQPGQW QQSGQGQHWY YPTSPQLSGQ GQRPGQWLQP GQGQQGYYPT SPQQPGQGQQ LGQWLQPGQG QQGYYPTSLQ QTGQGQQSGQ
GQQGYYSSYH VSVEHQAASL KVAKAQQLAA QLPAMCRLEG GDALSASQ

High molecular weight glutenin subunit 1Ay/Tu-e1 OS=Triticum urartu GN=Glu-A1-2 PE=4 SV=1 - [B5TM05_TRIUA]



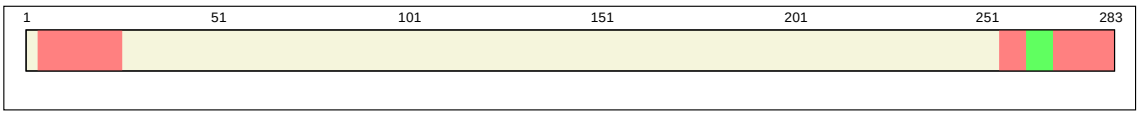
MAKRLVLFAT VVIGLVALTV AEGEASRQLQ CERELQESSL EACRLVVDQQ LAGRLPWSTG LQMRCCQQLR DISAKCRPDA VSQVARQYGQ TAVPPKGGSF
YSRETTPLQQ LQQGIFGGTS SQTVQGYYPS VISPQQGSYY PGQASPQQPG KWQELGQGQQ GYYPTSLQQP GQGQQGYYRT SLQQPGQGQQ GYYRTSLQQP
GQGQQIGQWQ QGYYPTSPQH PGQGQQPGQV QKIGQGQQPE KGQQLGQEQQ IGQGQQPEQG QQPGQGQQPG QGQQGYYPTS PQQPRQGQQP GQWQQPGQGQ
KGYYPTSLQQ PGQGQQGHYP ASQHQPGQGQ QGHHPASLQQ SGQGQQGHHP ASLQQPGQGK QTGQREQRQQ PGQGQQTGQG QQPEQEQQPG QGQQGYYPTY
PQQPGQGQQP EQWQQPGQGQ QRHYPASLQQ SGQGQQGHYP ASLQQPGQGQ PGQTQQPGQG QHPEQEEQPG QGQQGYYPTS PQQPGQGQQP GQGQQGHFPT
SGQAQQPGQG QQIGQAQQLG QGQQGYYPTS LQQPGQEQQS GQGQQLGQGH QPGQGQQSGQ EQQGYDSPYH VSVEQQAASP KVAKAHHPVA QLPTMCQMEG
GDALSASQ

Alpha gliadin OS=Triticum aestivum GN=gli-w7-104 PE=4 SV=1 - [A7LHB1_WHEAT]



MKTFLILALL AIVATTATTA VRVPVPQLQP QNPSQQQPQE QVPLVQQQQF LGQQQPFPPQ QPYPQPQPFP QQPYLQLQP FPQPQLPYSQ PQPFRPQQPY
PQPQPQYSQP QQPISQQQQQ QQQQQQQQQQ QQQILQQILQ QQMIPCRDVV QQHNIAHGR SQVLQQSRYQ LLQQLCCQQL WQIPEQSRCQ AIHNVVHAII
LHQQQQQQQQ QPLSQVCFQQ SQQQYPSGQG FQPSQQNPQ AQGSVQPQQL PQFEEIRNLA LETLPAMCNV YIPPYCTIAP VGIFGTN

Low molecular weight glutenin (Fragment) OS=Triticum aestivum PE=2 SV=1 - [Q5MFN4_WHEAT]



ETK**CIPGLER PWHLQPLPPQ QTFPR**EPLFS QEQQLFPQQP SFSQQQPPFW QQQPPFSQQQ PILPQQPPFS QQQQLVLPQQ PPFSQQQQPV LPPQQSPFPQ
QQQHQQLVQQ QIPFVHPSIL QQLNPCKVFL QQQCSPVAMP QSLARSQMLQ QSSCHVMQQQ CCQQLPQIPQ QSRYEAIRAI IYSIILQEQQ QVQGSIQTPQ
QQPQQLGQCV SQPQQQSQQQ LGQQPQQQQL AQGTFLQPHQ IAQLEVMTSI ALR**TLPTMCR VNVPLYRTTT SVPFGVGTGV GSY**

Table S2 - Protein identification according to the peptides sequenced using Proteome Discoverer 1.4 software.

PEPTIDE	RT (MIN)	MW (kDa)	PERCURSOR PROTEIN	
			CODE	NAME
BAND 1`				
QQQQQK	25.22	787.40886	R9XW75	Alpha-gliadin
TTTRVPFGVG	23.04	1034.5626	D6RVY7	LMW-Glutenin
TTTRVPFGVGTGVGG	25.08	1405.7448	Q8W3W5	LMW-Glutenin
TTTRVPFGVGTGVGGY	27.39	1568.8096	Q8W3W6	LMW-Glutenin
TLPTMCNVNVSLYR	32.71	1610.7651	B8XU58	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSR	40.83	3272.5433	B8XU58	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSR	40.83	3272.5433	D6RVY7	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSRYEAIR	23.72	3904.797	Q8W3W8	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSRYEAIR	23.72	3904.797	B8XU58	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSRYEAIR	23.72	3904.797	D6RVY7	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSR	40.83	3272.5433	D6RVY7	LMW-Glutenin
BAND 2´				
TITTTR	21.37	692.39818	C3W5R6	HMW-Glutenin
MSLQALR	22.11	818.45458	AVLB3	Avenin
TRATNKTR	22.27	947.52953	C3W5R6	HMW-Glutenin
TRATIRTR	33.03	974.57372	C3W5R6	HMW-Glutenin
VLPNFSAAAR	17.53	1045.5624	C3W5R6	HMW-Glutenin
TTNRTARTK	22.08	1048.5776	C3W5R6	HMW-Glutenin
VLPNFSASVR	33.64	1089.6247	C3W5R6	HMW-Glutenin
MARTGTRATR	39.39	1120.6329	C3W5R6	HMW-Glutenin
APFASIVAGIGGQ	33.63	1187.6421	B6DQB2	Gamma-Gliadin
AVPSFSAAARTR	22.2	1233.6434	C3W5R6	HMW-Glutenin
ALPSFSAAGRTR	22.2	1233.6434	C3W5R6	HMW-Glutenin
FSTTARTRATR	19.47	1267.6748	C3W5R6	HMW-Glutenin
GTFLQPHQIAR	19.65	1267.6926	B1A3G9	LMW-Glutenin

ATTALPSFSAATR	20.76	1293.6938	C3W5R6	HMW-Glutenin
VLPNFSAAARTR	19.3	1302.6999	C3W5R6	HMW-Glutenin
CQAIHNVAEAIR	25.97	1324.6997	AVLB3	Avenin
TRPTNRTSATTR	21.18	1361.6843	C3W5R6	HMW-Glutenin
IPEQSRYEAIR	17.56	1361.7157	B1A3B9	LMW-Glutenin
ATRVLSNFSTAAR	20.01	1393.7844	C3W5R6	HMW-Glutenin
TTTRVPFGVGTGVGG	24.75	1405.7474	Q8W3W5	LMW-Glutenin
TTRVLPNFSATPR	34.44	1459.7787	C3W5R6	HMW-Glutenin
TGATRVLSNFSAAAR	25.74	1521.8088	C3W5R6	HMW-Glutenin
TTTRVPFGVGTGVGGY	27.29	1568.8082	Q8W3W6	LMW-Glutenin
VLPNFSAAARTRATR	36.89	1630.8804	C3W5R6	HMW-Glutenin
ATRVLPNFSAAARTR	36.89	1630.8804	C3W5R6	HMW-Glutenin
ATRILPNFSAAARTR	23.85	1644.8869	C3W5R6	HMW-Glutenin
VLPNFSATPRTKATTR	25.88	1759.9577	C3W5R6	HMW-Glutenin
VFLQQQCSPVAMPQSLAR	28.1	2003.0038	B2Y2R3	LMW-Glutenin
VFLQQQCSPVAMPQSLAR	28.1	2003.0038	Q8W3W5	LMW-Glutenin
TLPTMCNVNVPLYRTTTR	32.1	2080.0408	Q8W3W6	LMW-Glutenin
TLPTMCNVNVPLYRTTTR	32.1	2080.0408	Q8W3W5	LMW-Glutenin
SLVLQTLPTMCNVYVPPECS	28.72	2194.0807	Q9FS77	Gamma-Gliadin
SLVLQTLATMCNVYVPPYCSTIR	33.37	2572.2689	B6DQB2	Gamma-Gliadin
RPLFQLVQGQGIIQPQQPAQLEVIR	32.03	2856.6246	Q1W676	Gamma-Gliadin
VFLQQQCSPVAMPQSLARSQMLQQR	37.77	2874.4141	B2Y2R3	LMW-Glutenin
QLGQGQQGYYPTSQQAGQGQQPGQGQQGR	31.32	3060.4195	X5CN16	HMW-Glutenin
SCHVMQQQCCQQLPQIPQQSRYEAIR	63.43	3102.5024	B2Y2R3	LMW-Glutenin
SMCNIYIPVQCPAPTTYNIPLVATYTGGAC	29.79	3161.5012	AVLB3	Avenin
TLPNMFNVYVRPDCSTINAPFASIVAGIGDQ	32.16	3310.686	B5ANT1	Gamma-Gliadin
NFLLQQCKPVSLVSSLWSMIWPQSDCQVMR	30.25	3523.6557	Q9FS77	Gamma-Gliadin
TFLVFALLAVVATSAIAQMETSCIPGLGETMGSNNHYNKK	63.97	4227.0124	B1A3G9	LMW-Glutenin
BAND 3´				
FTKMADPK	20.62	937.48668	Q0Q5D1	Protein Kinase
EPLTSETKK	38.71	1032.5788	Q0Q5D1	Protein Kinase
YWTHGGSLR	21.62	1076.4987	Q6RK62	Dof DNA-binding protein
VPFGVGTGVGGY	30.71	1109.5616	Q8W3W6	LMW-Glutenin
KNGASSSEPDK	22.65	1119.5242	Q0Q5D1	Protein Kinase

NGASSSEPDKK	22.65	1119.5242	Q0Q5D1	Protein Kinase
WFKNKIWR	34.97	1177.6784	Q0Q5D1	Protein Kinase
FICKLVAGVAGQ	23.18	1205.6895	B8XU43	Gamma-Gliadin
EISLPNSPKAR	25.46	1211.6401	Q0Q5D1	Protein Kinase
AKNKQLATLVK	51.07	1213.7491	Q0Q5D1	Protein Kinase
NVPIGGGCRKPK	40.64	1225.7111	Q6RK62	Dof DNA-binding protein
REAGQEASTSETK	24.02	1393.6143	Q0Q5D1	Protein Kinase
TTTRVPFGVGTGVGG	25.3	1405.7443	Q8W3W8	LMW-Glutenin
TTTRVPFGVGTGVGGY	27.64	1568.8078	Q8W3W6	LMW-Glutenin
MGCCCSWIRGIRTVK	28.29	1712.8183	Q0Q5D1	Protein Kinase
WRRSFCGWTSHEAK	27.95	1750.845	Q0Q5D1	Protein Kinase
GLEYLHEVANPPVVYR	23.65	1855.9735	Q0Q5D1	Protein Kinase
GSCACCCSWISGLCGRKK	17.69	1861.7847	Q0Q5D1	Protein Kinase
GYIPETMEVIAVKQLDK	23.51	1934.0246	Q0Q5D1	Protein Kinase
TLPTMCNVNVPLYRTTTR	32.46	2080.0389	Q8W3W6	LMW-Glutenin
TLPTMCNVNVPLYRTTTR	32.46	2080.0389	Q8W3W5	LMW-Glutenin
GFCGWTSHEAEEPSISETK	28.71	2094.9449	Q0Q5D1	Protein Kinase
GLEYLHEVANPPVVYRDLK	28.56	2212.111	Q0Q5D1	Protein Kinase
SFCGWTSHEAKEPLTSETKK	31.04	2266.1474	Q0Q5D1	Protein Kinase
YWTHGGSLRNVPIGGGCRKPK	30.34	2283.1492	Q6RK62	Dof DNA binding protein
QVVDQQLAGRLPWSTGLQMR	30.46	2283.1813	Q84TG6	HMW-Glutenin
IAVDAARGLEYLHEVANPPVVYR	48.25	2552.3274	Q0Q5D1	Protein Kinase
ILVYEYMPLGSLQDHLLDLTPK	28.75	2558.3108	Q0Q5D1	Protein Kinase
GQQTRQGQQLEQGQQPGQGQQTR	24.75	2566.2658	Q84TG6	HMW-Glutenin
DCYWINNGGSNPWQSLLNSSSLM	32.25	2586.1813	Q6RK62	Dof DNA binding protein
KWGRGFCGWTSHEAEEPSISETK	30.07	2622.2489	Q0Q5D1	Protein Kinase
WGRGFCGWTSHEAEEPSISETKK	30.07	2622.2489	Q0Q5D1	Protein Kinase
MEEVFPSNSKSKAGQMAGEAIAGAEK	27.84	2667.2477	Q6RK62	Dof DNA binding protein
MKIAVDAARGLEYLHEVANPPVVYR	53.36	2811.4449	Q0Q5D1	Protein Kinase
VAKVQQPATQLPIMCRMEGGDALSASQ	50.79	2829.4363	Q84TG6	HMW-Glutenin
SLELTTLPDCAKCMSDLTAPHQRFICK	36.48	3021.4121	B8XU42	Gamma-Gliadin
VFTHSQLSDATNSFSQENLLGEGGFGRVYK	30.41	3288.5307	Q0Q5D1	Protein Kinase

BAND 4´				
VPFGVGTGVGGY	30.22	1109.5619	Q8W3W6	LMW-Glutenin
APFASIVADIGGQ	34.86	1245.6486	Q94G97	Gamma-Gliadin
VIQEAKTLPPR	38.84	1251.7537	B0FHG5	Puroindoline
NYLLGRCSTMK	22.03	1285.6428	B0FHG5	Puroindoline
LNSCRNYLLGR	14.66	1308.6535	B0FHG5	Puroindoline
TLPTMCNVYVPR	19.2	1393.7303	Q1W676	Gamma-Gliadin
AQQLAAQLPAMCR	27.87	1400.6829	Q6R2V1	HMW-Glutenin
TTTRVPFGVGTGVGG	24.92	1405.7443	Q8W3W8	LMW-Glutenin
DFPVTWRWWK	27.88	1420.6769	B0FHG5	Puroindoline
DVVLQQHNIAHGR	16.26	1486.7858	F6M8E2	Alpha-Gliadin
TTTRVPFGVGTGVGGY	27.45	1568.8073	Q8W3W6	LMW-Glutenin
DVVLQQPNIAHASSK	20.33	1606.8556	I0IT55	Alpha/Beta Gliadin
TLPPRCNQGPPCDIP	29.66	1607.8047	B0FHG5	Puroindoline
ELKACQQVMDQQLR	27.79	1689.8023	Q6R2V1	HMW-Glutenin
GSVQPQQLPQFEEIR	26.96	1755.9036	R4VEK6	Gliadin
GGSFYPGETTPPQQLQQR	23.64	1990.964	Q6R2V1	HMW-Glutenin
TLPTMCNVNVPLYRTTTR	31.9	2080.044	Q8W3W8	LMW-Glutenin
TLPTMCNVNVPLYRTTTR	31.9	2080.044	Q8W3W6	LMW-Glutenin
CNIIQGSIQGDLSGIFGFQR	26.71	2153.1348	B0FHG5	Puroindoline
DISPECHPVVVSPVAGQYER	28.06	2182.0849	Q6R2V1	HMW-Glutenin
VIQEAKTLPPRCNQGPPCDIP	46.53	2276.0853	B0FHG5	Puroindoline
SLVLGTLPTMCNVFVPPECSTTK	30.39	2437.1856	Q94G97	Gamma-Gliadin
RPLFQLVQGQGIIQPQQPAQLEVIR	32.21	2856.6267	Q1W676	Gamma-Gliadin
SQMLQQSSCHVMQQQCCQQLPQIPQQSR	23.21	3272.4236	Q8W3W8	LMW-Glutenin
SQMLQQSSCHVMQQQCCQQLPQIPQQSR	23.21	3272.4236	Q8W3W6	LMW-Glutenin
ACQQVMDQQLRDISPECHPVVVSPVAGQYER	15.41	3482.5767	Q6R2V1	HMW-Glutenin
MKTFLILALLAIVATTATIAILQQILQQQLIPCR	53.59	3750.1245	F6M8E2	Alpha-Gliadin
GGCLELLGECCSQLGQMPPQCRCNIIQGSIQGDLSGIFGFQR	32.82	4456.1227	B0FHG5	Puroindoline
BAND 5`				
GSVQPQQLPQFEEIRNLALQTLPAMCNVYIPPYCTIAPFGIFGNT			RAVEK6	Gliadin
BAND 6`				
NYLLDR	21.04	793.4196	A6N860	Puroindoline
VNVPLYR	21.35	860.49816	Q5MFN4	LMW-Glutenin
LGGFLGIWR	33.22	1018.583	A8R0D1	Puroindoline
WWKWWK	22.52	1019.5404	A6N860	Puroindoline
LGQMPPQCR	18.99	1029.496	A6N860	Puroindoline
VPFGVGTGVGGY	21.35	1109.5373	B2Y2R3	

GGCEHEVREK	64.01	1143.5193	A8R0D1	Puroindoline
QLSQIAPQCR	32.13	1143.6148	A8R0D1	Puroindoline
GEVFKQLQRA	27.64	1175.6515	A8R0D1	Puroindoline
CCRQLESVSR	16.63	1180.5865	Q0Q5E3	LMW-Glutenin
LPWSTGLQMR	15.38	1188.61	B5TM05	HMW-Glutenin
APFASIVADIGGQ	34.74	1245.6467	Q94G97	Gamma-Gliadin
CDSIRRVIQGR	19.03	1302.6997	A8R0D1	Puroindoline
GEVREKPLLACR	18.61	1370.7563	Q0Q5E3	LMW-Glutenin
TLPTMCNVYVPR	18.96	1393.7314	Q1W676	Gamma-Gliadin
GEVREKPLHACR	20.72	1394.714	Q0Q5D9	Globulin
TTTRVPFGVGTGVGG	24.53	1405.7464	Q8W3X2	LMW-Glutenin
DFSVTWRWWK	15.7	1410.7088	A6N860	Puroindoline
DYVMERCFTMK	17.87	1422.6028	A8R0D1	Puroindoline
QSLTDAQCRGEVR	15.26	1462.7278	Q0Q5E3	LMW-Glutenin
VMRQQCCQQLAR	36.97	1463.6968	Q94G97	Gamma-Gliadin
ERCCRQLESVSR	18.54	1465.7347	Q0Q5E3	LMW-Glutenin
DYEQSMPPLGEGR	33.07	1478.6286	Q0Q5D9	Globulin
DYEQSMPPLGEGR	33.07	1478.6286	Q0Q5E3	LMW-Glutenin
DVVLQQHNIAHGR	15.93	1486.7853	A7LHB1	Alpha-Gliadin
TTTSVPFGVGTGVGSY	31.06	1529.7494	Q5MFN4	LMW-Glutenin
TTTRVPFGVGTGVGGY	27.02	1568.8069	B2Y2R3	LMW-Glutenin
DFPVTWPTKWWK	31.49	1590.783	A8R0D1	Puroindoline
ELKACQQVMDQQLR	27.8	1689.8034	SAU5H5	HMW-GS
QLSQIAPQCRCDSIR	24.49	1717.8376	A8R0D1	Puroindoline
EKCCKQLSQIAPQCR	26.99	1734.8139	A8R0D1	Puroindoline
AGEGAVGVPLFQAQWGAR	31.91	1813.9343	Q0Q5E3	LMW-Glutenin
AGEGAFGVPLFQAQSDAR	30.7	1820.8953	Q0Q5D9	Globulin
GGSFYPGETTPPQQLQQR	23.47	1990.9655	SAU5H5	HMW-GS
QLQCERELQESSLEACR	22.27	2021.8885	B5TM05	HMW-Glutenin
NYLLDRCSTMKDFSVTWR	27.42	2235.074	A6N860	Puroindoline
CRPDAVSQVARQYGQTAVPPK	24.95	2271.1408	B5TM05	HMW-Glutenin
WWKWWKGGCQELLGECCSR	31.21	2355.0342	A6N860	Puroindoline
LPWSTGLQMRCCQQLRDISAK	53.63	2434.1487	B5TM05	HMW-Glutenin
SLVLGTLPTMCNVFVPPECSTTK	31.15	2437.1768	Q94G97	Gamma-Gliadin
DYVMERCFTMKDFPVTWPTK	29.79	2494.1728	A8R0D1	Puroindoline
YQLLQQLCCQQLWQIPEQSR	55.05	2505.1959	A7LHB1	Alpha-Gliadin
CIPGLERPWHLQPLPPQQTFPR	26.24	2610.3138	Q5MFN4	LMW-Glutenin
YYPSVTCPQQVSYYPGQASPQRPGR	53.01	2829.4076	SAU5H5	HMW-GS
RPLFQLVQGQGIIQPQQPAQLEVIR	32.08	2856.631	Q1W676	Gamma-

				Gliadin
WWKGGCQELLGECCSRLGQMPPQCR	44.48	2865.3069	A6N860	Puroindoline
VFLQQQCSPVAMPQSLARSQMLQQR	37.69	2874.398	B2Y2R3	LMW-Glutenin
DISPECHPVVVSPVAGQYEQQIVVPPK	61	2915.4576	SAU5H5	HMW-GS
NYLLQQCNPVSLVSSLVSMILPRSDCK	50.34	3007.5296	Q94G97	Gamma-Gliadin
TLPTMCNVNVPLYRTTTRVPFGVGTGVGG	47.89	3007.5525	Q8W3X2	LMW-Glutenin
IFALLAVAATSAIAQMENSHIPGLERPSR	49.25	3063.7068	Q8W3X2	LMW-Glutenin
SCHVMQQQCCQQLPQIPQQSRYEAIR	12.12	3102.4927	B2Y2R3	LMW-Glutenin
TLPTMCRVNVPLYRTTTSVPFGVGTGVGSY	47.7	3173.5213	Q5MFN4	LMW-Glutenin
QEVQGGQYGSETGGSQQQQQGGGYHGVTVGRGGQRQGQVLCHK	34	4456.1061	Q0Q5E3	Globulin
QPERGCSGESTAEQRQEVQGGQYGSETGGSQQQGGGYHGVTVGR	33.67	4524.1	Q0Q5D9	Globulin
DISPECHPVVVSPVAGQYEQQIVVPPKGGSFYPGETTPPQQLQQR	21.03	4887.3018	SAU5H5	HMW-GS